Fusion of Large Continuously Collected Data Sources: Understanding Travel Demand Trends and Measuring Transport Project Impacts

Gregory D. Erhardt

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Centre for Advanced Spatial Analysis

University College London

July 29, 2016

I, Gregory D. Erhardt, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

This research combines several large, continuously collected data sets to understand recent travel demand trends in San Francisco, and it develops a tool for measuring transport project impacts.

Because they are continuously collected, these data provide an opportunity to measure change in a way that is not available in traditional, cross-sectional travel surveys. The data used are from San Francisco and cover performance of the transit system and associated factors expected to drive transit demand.

This study employs a two stage methodology to derive insight from these data. First, a performance monitoring tool is developed to process the raw data and report meaningful performance indicators. This tool encapsulates the necessary data cleaning functionality, and manages a multi-stage data expansion process to ensure that data are representative of the system as a whole. Second, time series models of transit ridership are estimated from the outputs of the performance monitoring tool. These time series models provide a means of quantifying the portion of the ridership changes due to service changes versus background factors, such as employment growth.

The estimated models are applied to understand the drivers of recent ridership trends in the San Francisco, where ridership on the San Francisco Municipal Railway (MUNI) bus system remains flat in spite of population and employment growth, while ridership on the Bay Area Rapid Transit (BART) system grows faster than employment. In addition, the models are applied to several planning case studies, including both ex-post ridership evaluations and short-term forecasting applications.

The outcome of this research is to establish and test a tool to facilitate the use of passively collected data for retrospective travel demand analyses. It provides insight into the effects of transport projects, and lays the groundwork for a future studies that further our ability to observe and understand travel behaviour.

Acknowledgements

Thank you to my PhD supervisors, Mike Batty and Elsa Arcaute, who have been excellent mentors and helped guide this work to completion. Your feedback has always been appreciated, as have your stories. Whatever problem I show up with, you can either point me in the right direction, or point me to the right person to talk to, and you never hesitate to do so.

The first year of this research was funded by the San Francisco County Transportation Authority. Thank you both to the Authority and to Elizabeth Sall who made the arrangements and engaged in some early thinking about the direction of the research. Subsequent funding was provided by the University College London Graduate Research Scholarship and Overseas Research Scholarship. Thank you to the university and to the donors for both scholarships.

Data for this study was provided by the San Francisco County Transportation Authority, the San Francisco Municipal Transportation Agency, the San Francisco Planning Department, the Metropolitan Transportation Commission, Bay Area Rapid Transit, Streetlight Data, the Massachusetts Institute of Technology, Stamen Design and to several additional agencies whose data are widely available for download. Thank you to staff from those agencies, especially to Stephen Granger-Bevan and David Ory who facilitated the obfuscation and release of the Clipper Card data.

Oliver Lock wrote his master's thesis using the outputs of the tool developed here, and created a complementary data visualisation tool. His engagement and testing have made this work better. We co-authored papers as a result, as described in Appendix A.

Thank you to my colleagues at the Centre for Advanced Spatial Analysis and at RAND Europe for welcoming me into your communities, for the opportunity to learn from you, and for many engaging modelling discussions.

During the viva examination, Peter Jones and Andrew Daly provided constructive feedback that improved the quality of this thesis.

The themes contained in this thesis reflect the influence of a number of mentors I have had throughout my career. Gordon Schultz gave me a remarkable amount of attention as a co-op student. Frank Koppelman instilled an appreciation for the theoretical foundations of modelling. Jeff May and Erik Sabina taught me what policy makers actually care about. Bill Davidson and Jim Ryan sparked my interest in before-and-after studies during our meetings at spring training. Joel Freedman taught me about everything from modelling, to programming, to project management, and was a tremendous resource throughout. Rick Donnelly fought battles, some of which I may never know about, to help me through some challenging projects with our skins and our sense of humour intact.

Thank you to my boys, Ansel and Ezra, for continuing to think that transportation is fun, loving train rides, and indulging me when I tell you that getting stuck in a traffic jam is research.

Most of all, thank you to my wonderful wife, Andrea. Your partnership and support have been invaluable, especially these last few months as I have leaned on you more heavily. I am lucky to have you to talk through my ideas with, edit my applications (especially to make sure that I correctly spell the name of the institution to which I am applying) and teach me how to be an academic. It is because of your appointment at Cambridge that I started a PhD in the first place, and it has been a great opportunity. It has been a pleasure following you around the world, and I very much look forward to our next stop. I am glad I got off the L that night...

Contents

Contents 14

List of Figures 18

List of Tables 21

Nomenclature 23

1 Introduction 27

1.1 Background 27

1.1.1 Problems with Transport Demand Forecasting Accuracy . . 27

1.1.2 Ex-post Evaluation of Transport Projects 32

1.1.3 Current State of the Field 34

1.1.4 The Emerging Opportunities of Big Data 35

1.2 Research Questions and Overall Approach 38

1.3 Available Data 40

1.3.1 Transit Data 40

1.3.2 Highway Data 41

1.3.3 Data Sources for Demand Drivers 42

1.3.4 Other Related Data 44

1.4 Data Fusion and Performance Monitoring 45

1.4.1 Key Design Goals 46

1.4.2 Data Flow 47

1.5 Analysis of Contributing Factors 49

1.5.1 Confounding Factors in Travel Demand 49

1.5.2 Methods for Controlling Confounding Factors 50

1.5.3 Selected Approach 52
1.6 Thesis Structure and Contribution 53
1.6.1 Structure 54
1.6.2 Research Contribution 55

2 Transit Smart Card Data Evaluation 57
2.1 Introduction 57
2.2 Context 59
2.3 Data 62
2.3.1 Clipper Card Data 62
2.3.2 Data Processing and Weighting 64
2.3.3 Other Supporting Data 65
2.4 Data Evaluation 67
2.4.1 System and Data Characteristics 67
2.4.2 Penetration Rates 69
2.4.3 Geographic Distribution 70
2.4.4 Temporal Distribution 71
2.4.5 Transfer Rates 71
2.4.6 Stability over Time 74
2.5 Estimated Model Correction Factors 75
2.5.1 Estimated Models of Clipper Card Use 75
2.5.2 Application of Correction Factors 79
2.6 Conclusions 82
2.6.1 Data Evaluation 82
2.6.2 Proposed Correction Factors 84

3 Transit Data Fusion 87
3.1 Introduction 87
3.2 Data Sources 89
3.3 Methodology 93
3.3.1 Cleaning Individual Data Sets 93

3.3.2 Data Expansion and Weighting 94
3.3.3 Percent of Trips Observed 101
3.3.4 Software Environment 104
3.4 Sample Results 105
3.5 Conclusions and Future Development 114

4 Time Series Model Estimation 117

4.1 Introduction 117
4.2 Data 123
4.2.1 MUNI 123
4.2.2 Other Transit Modes 127
4.2.3 Expected Drivers of Demand 131
4.2.4 Commute Mode Shares 136
4.2.5 Divergence of Ridership and Employment 138
4.3 Methodology 144
4.3.1 Measuring Autocorrelation 145
4.3.2 Stationarity and Differencing 147
4.3.3 ARIMA Models 148
4.3.4 Regression Models 151
4.3.5 Regression Models with ARIMA Errors 152
4.3.6 Lagged Variables 155
4.3.7 Goodness of Fit Measures 157
4.4 Model Results 158
4.4.1 Stationarity and Differencing 158
4.4.2 ARIMA Models 166
4.4.3 Regression Models 171
4.4.4 Regression Models with ARIMA Errors 182
4.4.5 Comparison of Results 193
4.5 Conclusions 198

5 Understanding Ridership Trends 201

5.1 Introduction 201
5.2 Literature Review 205
5.3 Working Hypotheses 209
5.3.1 Different Level-of-Service Trends 209
5.3.2 Suburban Growth Exceeds Urban Growth 211
5.3.3 MUNI Trips Shift to Rail 214
5.3.4 Increased Cost of Car Travel 214
5.3.5 Transit Riders are Priced out of San Francisco 219
5.3.6 The “Uber Effect” 221
5.4 Understanding Demand Changes 227
5.4.1 Model Elasticities 228
5.4.2 Calculating the Factors Contributing to Ridership Changes . 231
5.4.3 Changes in MUNI Ridership 233
5.4.4 Changes in BART Ridership 238
5.4.5 Comparison for 2009 to 2013 Period 250
5.5 Hypotheses Revisited 253
5.5.1 Different Level-of-Service Trends 253
5.5.2 Suburban Growth Exceeds Urban Growth 254
5.5.3 MUNI Trips Shift to Rail 254
5.5.4 Increased Cost of Car Travel 255
5.5.5 Transit Riders are Priced out of San Francisco 255
5.5.6 The “Uber Effect” 255
5.6 The Effect of an Ageing Population 257
5.6.1 Model Estimation 258
5.6.2 Ridership Changes 261
5.7 Conclusions 264
5.7.1 What is Driving the Ridership Divergence? 264
5.7.2 Future Research 266
5.7.3 Implications for Travel Forecasting 266

6 Project-Level Applications 269

6.1 Ex-Post Applications 271

6.1.1 2010 MUNI Service Cuts 273

6.1.2 2009 BART Service Cuts 277

6.1.3 2003 BART SFO Extension 280

6.2 Forecasting Applications 285

6.2.1 MUNI 5L Fulton Pilot Project 287

6.2.2 BART Perks Pilot Project 290

6.3 Conclusions 295

6.3.1 Ex-Post Applications 295

6.3.2 Forecasting Applications 296

6.3.3 Integration into Practice 297

6.3.4 Future Research 298

7 Conclusions 299

7.1 Research Findings 299

7.1.1 Transit Smart Card Data Evaluation 300

7.1.2 Transit Data Fusion 302

7.1.3 Time Series Model Estimation 303

7.1.4 Understanding Ridership Trends 304

7.1.5 Project-Level Applications 305

7.2 Future Research Directions 306

7.2.1 Incorporating Highway Measures 306

7.2.2 Combining with Rolling Household Travel Surveys 307

7.2.3 Panel Data Models 308

7.2.4 Integration into Practice 308

7.3 Lessons Learned 309

7.3.1 Plan Ahead for Data Archival 309

7.3.2 Big Data is Not a Replacement for Survey Data 310

7.3.3 Data Access Constraints are a Barrier to Research 311

7.3.4 Strategies for Data Privacy 311

7.3.5 The Role of Technology Companies 312
7.4 Implications for Travel Forecasting 314

References **316**

Appendix A Published Works **351**

Appendix B Enumeration of Full Monthly Time Series **355**

Appendix C Model Application Formulas **397**

List of Figures

1.1 Ratio of actual to forecast ridership for US urban rail transit projects from three study cohorts [3] 29
1.2 Ratio of actual to forecast traffic for toll roads [4] 30
1.3 Data fusion tool design 48

2.1 Bay Area transit operators [142] 61
2.2 Trips between BART station pairs estimated from Clipper Card transactions and fare gate totals 72
2.3 Time-of-day distribution from Clipper transactions and from AVL/APC data on MUNI buses 73
2.4 Trend in Clipper tag-ons on BART and MUNI, compared to total boardings 75

3.1 Data processing flow 93
3.2 Sample transit performance change map 111
3.3 Sample route profile 112
3.4 Sample trend plot 113

4.1 MUNI and BART riders per employee 139
4.2 Trends in MUNI ridership and San Francisco employment 140
4.3 Decomposition of MUNI riders per employee 141
4.4 Trends in BART ridership and 4-county employment 142
4.5 Decomposition of BART riders per employee 143
4.6 Example ACF and PACF plots 146
4.7 Example of spurious correlation [198] 153
4.8 Differencing options for MUNI riders 159

4.9 Autocorrelation of MUNI riders 161
4.10 Autocorrelation of seasonal difference in MUNI riders 161
4.11 Autocorrelation of second difference in MUNI riders 162
4.12 Differencing options for BART riders 163
4.13 Autocorrelation of BART riders 164
4.14 Autocorrelation of seasonal difference in BART riders 164
4.15 Autocorrelation of second difference in BART riders 165
4.16 Residual autocorrelation from MUNI ARIMA model 167
4.17 MUNI boardings, observed vs. ARIMA model 168
4.18 Change in MUNI boardings, observed vs. ARIMA model 168
4.19 Residual autocorrelation from BART ARIMA model 170
4.20 BART boardings, observed vs. ARIMA model 170
4.21 Change in BART boardings, observed vs. ARIMA model 171
4.22 Residual autocorrelation from MUNI regression model 176
4.23 MUNI boardings, observed vs. regression model 176
4.24 Change in MUNI boardings, observed vs. regression model 177
4.25 Residual autocorrelation from BART regression model 180
4.26 BART boardings, observed vs. regression model 181
4.27 Change in BART boardings, observed vs. regression model 181
4.28 Residual autocorrelation from MUNI RegARIMA model 184
4.29 MUNI boardings, observed vs. RegARIMA model 185
4.30 Change in MUNI boardings, observed vs. RegARIMA model . . . 185
4.31 Residual autocorrelation from MUNI RegARIMA model with constrained employment term 187
4.32 MUNI boardings, observed vs. RegARIMA model with constrained employment term 187
4.33 Change in MUNI boardings, observed vs. RegARIMA model with constrained employment term 188
4.34 Residual autocorrelation from BART RegARIMA model 192
4.35 BART boardings, observed vs. RegARIMA model 192

4.36 Change in BART boardings, observed vs. RegARIMA model . . . 193

5.1 Bay Area transit system [142] 202
5.2 Trend in Bay Area transit ridership 203
5.3 MUNI and BART service miles 211
5.4 MUNI and BART fares 212
5.5 Employment in San Francisco and 4-county area 213
5.6 Population in San Francisco and 4-county area 215
5.7 Workers in San Francisco and 4-county area 216
5.8 Service miles on SFMTA submodes 217
5.9 Ridership on SFMTA submodes 218
5.10 Car travel costs 220
5.11 Households in San Francisco by income group 222
5.12 Workers living in San Francisco by earnings 223
5.13 Transit mode share by earnings for workers living in San Francisco 224
5.14 Walk, bike, taxi and other commute mode shares for workers living in San Francisco 225
5.15 William Playfair's plots of English exports and imports to and from Denmark and Norway [229] 232
5.16 Effect of changes in service miles on MUNI ridership, vs. Sep 2009 234
5.17 Effect of changes in rail service on MUNI ridership, vs. Sep 2009 . 235
5.18 Effect of changes in runspeed on MUNI ridership, vs. Sep 2009 . . 236
5.19 Effect of changes in employment on MUNI ridership, vs. Sep 2009 237
5.20 Effect of unexplained trend on MUNI ridership, vs. Sep 2009 . . . 237
5.21 Effect of residuals on MUNI ridership, vs. Sep 2009 238
5.22 Effect of changes in service miles on BART ridership, vs. Jan 2001 239
5.23 Effect of changes in service miles on BART ridership, vs. Sep 2009 240
5.24 Effect of changes in stations on BART ridership, vs. Jan 2001 . . . 240
5.25 Effect of changes in stations on BART ridership, vs. Sep 2009 . . . 241
5.26 Effect of changes in employment on BART ridership, vs. Jan 2001 . 242
5.27 Effect of changes in employment on BART ridership, vs. Sep 2009 242

5.28 Effect of changes in percent of employment in San Francisco on BART ridership, vs. Jan 2001 243
5.29 Effect of changes in percent of employment in San Francisco on BART ridership, vs. Sep 2009 244
5.30 Effect of changes in fare on BART ridership, vs. Jan 2001 244
5.31 Effect of changes in fare on BART ridership, vs. Sep 2009 245
5.32 Effect of changes in car fuel cost on BART ridership, vs. Jan 2001 . 245
5.33 Effect of changes in car fuel cost on BART ridership, vs. Sep 2009 . 246
5.34 Effect of strikes on BART ridership, vs. Jan 2001 247
5.35 Effect of strikes on BART ridership, vs. Sep 2009 247
5.36 Effect of unexplained trend on BART ridership, vs. Jan 2001 248
5.37 Effect of unexplained trend on BART ridership, vs. Sep 2009 249
5.38 Effect of residuals on BART ridership, vs. Jan 2001 249
5.39 Effect of residuals on BART ridership, vs. Sep 2009 250
5.40 Unexplained trend in MUNI ridership vs. walk, bike, taxi and other commuters 257
5.41 Residual autocorrelation from MUNI RegARIMA model with constrained employment term and age 260
5.42 MUNI boardings, observed vs. RegARIMA model with constrained employment term and age 260
5.43 Change in MUNI boardings, observed vs. RegARIMA model with constrained employment term and age 261

6.1 Overview of service for 5L Fulton pilot project [254] 288
6.2 MUNI ridership forecast 289
6.3 BART ridership forecast 293

List of Tables

2.1 Clipper data dictionary . . . 63
2.2 Transit system characteristics . . . 68
2.3 Estimated Clipper Card penetration rates . . . 70
2.4 Transfers per linked trip, from Clipper transactions . . . 73
2.5 Transfers per linked trip, from onboard surveys . . . 74
2.6 Estimated models of Clipper Card use . . . 77
2.7 Correction factors applied to onboard survey data, totals by operator 80
2.8 Correction factors applied to onboard survey data, shares by category 81

3.1 Data dictionary for expanded and weighted trip-stop Datastore . . . 96
3.2 Trip weighting example . . . 99
3.3 Percent of trips observed on each route on weekdays in July 2010 . . 102
3.4 Sample transit performance summary report . . . 106

4.1 Trends from MUNI performance report . . . 124
4.2 Annual differences from MUNI performance report . . . 125
4.3 Trends for other transit modes . . . 129
4.4 Annual differences other transit modes . . . 130
4.5 Expected drivers of demand trends . . . 132
4.6 Expected drivers of demand annual differences . . . 133
4.7 Mode share trends . . . 137
4.8 Mode share annual differences . . . 137
4.9 Example of distributed lag calculations . . . 156
4.10 ARIMA model of MUNI boardings . . . 166
4.11 ARIMA model of BART boardings . . . 169

4.12 Regression model of MUNI boardings 172
4.13 Regression model of BART boardings 178
4.14 RegARIMA model of MUNI boardings 182
4.15 RegARIMA model of MUNI boardings with constrained employment term 186
4.16 RegARIMA model of BART boardings 190
4.17 Three models of MUNI boardings 194
4.18 Comparison of MUNI RegARIMA Models with and without employment constraint 194
4.19 Three models of BART boardings 196
4.20 Comparison of BART regression and RegARIMA models with equivalent specifications 197
4.21 Comparison of MUNI and BART RegARIMA Models 198

5.1 RegARIMA Models of MUNI boardings with constrained employment term 227
5.2 RegARIMA models of BART boardings 228
5.3 Elasticities from MUNI RegARIMA model with unconstrained employment term 229
5.4 Elasticities from MUNI RegARIMA model with constrained employment term 230
5.5 Elasticities from BART RegARIMA model 231
5.6 Contributions to change in MUNI ridership: Sep 2009 to Sep 2013 . 251
5.7 Contributions to change in BART ridership: Sep 2009 to Sep 2013 . 251
5.8 Comparison of factors contributing to changes in MUNI vs. BART riders: Sep 2009 to Sep 2013 253
5.9 RegARIMA model of MUNI boardings with constrained employment term and age 259
5.10 Comparison of MUNI RegARIMA Models with and without age term 262
5.11 Contributions to change in MUNI ridership, considering age effect: Sep 2009 to Sep 2013 263

5.12 Comparison of contributions to change in MUNI ridership, with and without age effect: Sep 2009 to Sep 2013 263

6.1 Estimated effect of MUNI 2010 service cuts 274
6.2 Before and after calculations for MUNI 2010 service cuts 276
6.3 Elasticity calculations for MUNI 2010 service cuts 276
6.4 Contributions to change in MUNI ridership: Aug 2009 to Aug 2010 277
6.5 Estimated effect of BART 2009 service cuts 279
6.6 Before and after calculations for BART 2009 service cuts and fare increase 279
6.7 Elasticity calculations for BART 2009 service cuts and fare increase 280
6.8 Contributions to change in BART ridership: Feb 2009 to Feb 2011 . 281
6.9 Estimated effect of BART SFO extension 283
6.10 Before and after calculations for BART SFO extension 284
6.11 Elasticity calculations for BART SFO extension 284
6.12 Contributions to change in BART ridership: Feb 2003 to Feb 2005 . 285
6.13 Baseline forecast of MUNI ridership 290
6.14 Validation forecast of BART ridership 294
6.15 Baseline forecast of BART ridership 294

Nomenclature

AC Transit Alameda County Transit.

ACF Autocorrelation Function.

ACS American Community Survey.

AIC Akaike's Information Criterion.

AICc Corrected Akaike's Information Criterion.

APC Automated Passenger Counter.

API Application Program Interface.

AR Autoregressive.

ARIMA Autoregressive Integrated Moving Average.

ATR Automated Traffic Recorder.

AVL Automated Vehicle Location.

BART Bay Area Rapid Transit.

BATA Bay Area Toll Authority.

BLS Bureau of Labor Statistics.

BTS Bureau of Transportation Statistics.

Caltrans California Department of Transportation.

CBD Central Business District.

CHTS California Household Travel Survey.

CPI Consumer Price Index.

CTA Chicago Transit Authority.

DTA Dynamic Traffic Assignment.

EIA US Energy Information Administration.

FTA Federal Transit Administration.

GNU GNU's Not Unix!.

GPS Global Positioning System.

GTFS General Transit Feed Specification.

HDF5 Hierarchical Data Format 5.

IRS Internal Revenue Service.

ITS Intelligent Transportation Systems.

LEHD Longitudinal Employer-Household Dynamics.

LODES LEHD Origin-Destination Employment Statistics.

LRV Light Rail Vehicle.

MLE Maximum Likelihood Estimation.

MSA Metropolitan Statistical Area.

MTA Metropolitan Transit Authority.

MTC Metropolitan Transportation Commission.

MUNI San Francisco Municipal Railway.

NCHRP National Cooperative Highway Research Program.

NOAA National Oceanic and Atmospheric Administration.

NTS UK National Travel Survey.

OD Origin-Destination.

OLS Ordinary Least Squares.

PACF Partial Autocorrelation Function.

PeMS California Performance Monitoring System.

PII Personally Identifiable Information.

QCEW Quarterly Census of Employment and Wages.

RAC Residence Area Characteristics.

RegARIMA Regression with ARIMA Errors.

RMSE Root Mean Square Error.

San Mateo County Transit SamTrans.

SARIMA Seasonal Autoregressive Integrated Moving Average.

SF-CHAMP San Francisco Chained Activity Modeling Process.

SFCTA San Francisco County Transportation Authority.

SFMTA San Francisco Municipal Transportation Agency.

SFO San Francisco International Airport.

SPUR San Francisco Bay Area Planning and Urban Research Association.

TCQSM Transit Capacity and Quality of Service Manual.

TCRP Transit Cooperative Research Program.

TCSC Time Series Cross Section.

TRB Transportation Research Board.

TSDC Transportation Secure Data Center.

TTI Texas A&M Transportation Institute.

VMT Vehicle Miles Travelled.

VTA Valley Transportation Authority.

WAC Workplace Area Characteristics.

Chapter 1

Introduction

This research involves developing a prototype tool for the fusion of several large, continuously collected data sources that monitor both travel demand trends and the expected drivers of those trends. It explores potential biases and limitations of such data, and applies the outputs of the tool to gain insight into the reasons for recent changes in transit demand on two systems in the San Francisco Bay Area.

1.1 Background

This section describes the current state of the travel forecasting field, and the opportunities that exist for advancement through the use of Big Data.

1.1.1 Problems with Transport Demand Forecasting Accuracy

Spending on transport infrastructure in the United Kingdom (UK) exceeds £10 billion annually, while spending in the United States (US) exceeds $70 billion annually [1]. Travel demand forecasts play a central role in selecting which projects get built, and justifying the investment in those projects.

Unfortunately, forecasts for major transport infrastructure projects are not always accurate, and have a tendency for optimism bias. This was infamously documented 25 years ago when Pickrell [2] showed that actual ridership on US urban rail projects was 68% lower than forecast ridership. The problem has continued, as can be observed in Figure 1.1, which shows the ratio of actual to forecast ridership from the 1990 Pickrell report (Figure 1.1a), from a 2003 update to the study (Figure 1.1b), and from a 2007 update (Figure 1.1c). On these charts, perfectly accurate forecasts would plot on the x-axis at 100%. An unbiased set of forecasts would have a roughly equal number of forecasts too high and too low, by roughly equal amounts. The black line on each chart indicates what would be expected if the ratio followed a normal distribution with a mean of 100% and the observed standard deviation. While the more recently completed projects show improvement, the bias remains, with actual ridership in the 2003 study projects 31% lower than forecast, and actual ridership in the 2007 study projects 25% lower than forecast [3].

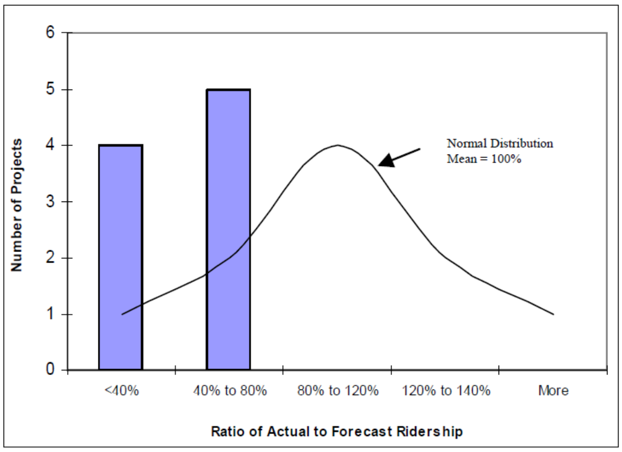
Forecasting errors are not limited to rail projects. Figure 1.2 shows the ratio of actual to forecast traffic for a global sample of toll roads, with the actual traffic 23% lower than forecast on average [4]. This has recently been problematic in Australia, where several privately financed toll roads have failed to achieve the forecast traffic and revenue. In response, investors have filed a series of high-profile lawsuits against the engineering firms responsible for the forecasts, seeking damages of up to A$1.6 billion [5, 6, 7].

Interestingly, untolled road projects do not show the same bias, with actual traffic typically exceeding forecast traffic by 3-11% [8]. Several other studies have documented accuracy issues for all types of studies, and found that even when there is no bias, actual travel demand shows substantial variance when compared to forecast demand [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19].

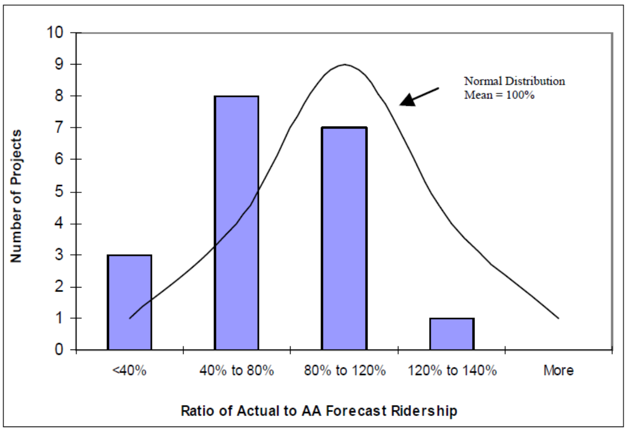
Of these studies, several focus on US urban rail projects [9, 11, 12, 16]. These are large projects within urban areas, often costing hundreds of millions of dollars. As a matter of policy, they receive the majority of their capital funding as a federal grant, awarded on a competitive basis with a key factor being the forecast ridership (thus providing an incentive for high forecasts). Figure 1.1 shows the actual versus forecast ridership comparisons for these projects.

A second body of work, focuses on toll road projects [4, 18], with the key results shown in Figure 1.2. These projects often involve private financing, either through an equity stake or through the issuance of bonds. Again, these are typically large scale projects, often, but not always within urban areas.

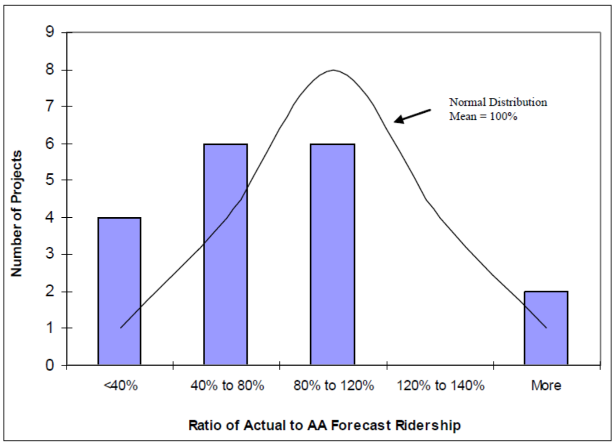
A third set of work is targeted at the analysis of "mega-projects" [10, 13, 14,



(a) *1990 Study Projects*



(b) *2003 Study Projects (AA=Alternatives Analysis)*



(c) *2007 Study Projects (AA=Alternatives Analysis)*

Figure 1.1: Ratio of actual to forecast ridership for US urban rail transit projects from three study cohorts [3]

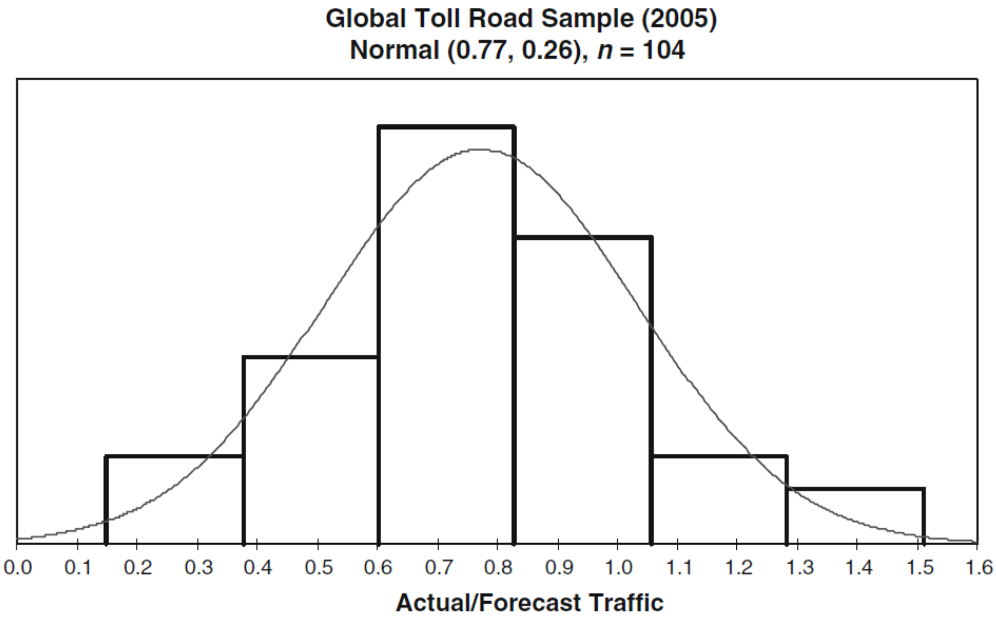


Figure 1.2: Ratio of actual to forecast traffic for toll roads [4]

15]. Often these are large-scale tunnel or bridge projects (either for road or rail) that cost billions of dollars, pounds or euros. The main conclusion here is that "the quality of demand forecasts is often poor, especially for rail projects" [15].

Understandably given the difference in the stakes involved, there has been somewhat less attention given to more routine projects. Hartgen [19] notes that there are a number of studies of traveller response to smaller projects, but a general lack of studies specifically of demand forecast accuracy for such projects. The Post-Opening Project Evaluations of Major Schemes (POPE) [20] does fill this role for highway projects in the UK, finding that 65% of schemes evaluated were within 15% of the forecast traffic volume, although with substantial variation between schemes. Parthasarathi and Levinson [17] evaluated 108 untolled roadway projects in Minnesota. They found that on average, actual traffic was higher than forecast traffic, although this varied by facility type. The forecasts tended to underestimate freeway traffic but overestimate lower class facilities.

A number of hypotheses have been put forward to explain the differences between actual and forecast travel demand. These explanations can be broadly grouped into three categories: technical problems, optimism bias, and selection bias.

Technical problems cover a wide variety of issues related to the models, the inputs, and the operation of those models. Specific examples cited include errors in forecasting exogenous variables such as fuel price or economic growth rates, inaccurate land-use projections, changes to the design or operation of the project after the forecasts are completed, assumptions about the competing facilities, and changes to underlying travel behaviour [21, 14, 15, 3, 16, 17, 19].

Another potential source of forecasting error identified by some is optimism bias, whereby those involved in developing planning forecasts make overly optimistic assumptions either due to unintentional psychological factors or due to political and institutional structures that incentivise such assumptions [9, 15, 16, 17, 18, 19]. Flyvbjerg [14] goes further to argue that such errors are the result of deliberate manipulation on the part of actors promoting the projects to overstate the project benefits.

Eliasson and Fosgerau [22] introduce an alternative explanation, that of selection bias. Selection bias occurs because those projects with higher forecast demand are more likely to be selected for construction. They show that in any planning system where ex ante forecasts affect which projects are selected, the projects that are ultimately built will be more likely to have higher than actual forecasts, even when the underlying forecasts for all potential projects are unbiased.

In their recent review of demand forecast accuracy, Nicolaisen and Driscoll [8] note that despite the multitude of *possible* explanations, there is little convincing evidence for the actual causes of forecast inaccuracy. To meet this need, they recommend mandatory ex-post evaluation schemes to provide evidence for such analysis.

The lack of evidence about the causes of travel forecasting error has not stopped authors from proposing solutions.

Flyvbjerg [14] recommends two possible solutions. The first is to shift to a method called reference-class forecasting, which involves developing project forecasts through a comparison to other, similar, projects. Flyvbjerg's second recommendation is to change the institutional incentives structure used to fund major projects towards more transparency and accountability. It is noteworthy that these recommendations rely on the availability of ex-post project evaluations, which would serve both to build a body of reference cases, and to make the performance of both the project and the forecasters transparent.

Hartgen [19] proposes either a "hubris" approach involving a large-scale investment in better methods, or a "humility" approach which de-emphasises the significance of travel forecasts by performing additional risk analysis, incorporating scenario planning, or shortening planning horizons. Again, a body of ex-post evaluations would enhance Hartgen's recommendations, providing information on the performance of different methods and on the demand risks commonly observed in projects.

1.1.2 Ex-post Evaluation of Transport Projects

Some ex-post evaluation schemes have become institutionalised in recent years, notably the evaluation of rail projects in the US and highway projects in the UK. In the US, since 2006, the Federal Transit Administration (FTA) is required to submit a report to congress evaluating the forecasting and cost accuracy of federally funded rail projects after their opening [23, 24, 12, 3, 25, 26, 27, 28, 29]. In the UK, since 2002, the Department for Transport (DfT) has evaluated 75 major highway schemes after their opening through their Post-Opening Project Evaluations of Major Schemes (POPE) [20]. Beyond these important examples, ex-post evaluation of transport projects does not appear to be either systematic or widespread [30, 31]. There is, however, an increased interest in the empirical evaluation of transport projects, motivated in part by recent US federal transportation legislation that requires states to establish performance-based planning programs [32].

Those authors that have conducted ex-post evaluations of transport projects for specific cases often note that assembling the necessary data is the component of the study requiring the largest effort, and an obstacle to further work [33, 17, 34]. Nicolaisen and Driscoll [8] echo this sentiment, stating that:

> "In general, lack of data availability has been an important obstacle in

> all reviewed studies and should be a key focus area of future research and practice on the use of demand forecasts as decision support. The lack of data access makes it difficult to perform more elaborate statistical analyses on exogenous variables, cover larger network effects, evaluate demand over time, and track changes in land use, project design or service levels."

Ex-post evaluations of transport investments are also faced with a number of methodological challenges, derived largely from the difficulty of separating treatment effects from other confounding factors.

Randomised control trials are often held up as the "gold standard" for treatment effects, particularly in the health sciences [35]. While such experiments are possible in transport, as with the Metropolitan Transit Authority (MTA) experiments in Boston conducted in the early 1960s [36], they remain rare [37]. There is a good reason that such studies are rare: transport infrastructure is expensive, and it would be difficult to justify hundreds of millions of pounds building a randomly selected road in the name of conducting an experiment. That is the purpose of building models in the first place: to estimate the effects of a project without having to pay for its construction. Axelrod [38] discusses the role of models as a "third way of doing science", in contrast to experiments and observational studies, and in practice, models do provide a means for conducting a simulated experiment where everything in the model is held constant except for the project itself. However, in this case, our goal is partially to understand whether the response of our models is correct, and we are left to conduct some form of empirical analysis to do so.

The Transit Cooperative Research Program (TCRP) Report 95 [39] is perhaps the most comprehensive effort to empirically measure the response to transport system changes, updating earlier work on the same topic [40, 41]. Between 2003 and 2013, TCRP 95 was published as sixteen stand-alone chapters [39, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56] focusing on different types of transport system changes, with three additional chapters yet to be completed. In compiling the available evidence on each topic from a range of existing studies, it confronts the issues in obtaining that evidence, and groups those issues into three primary areas: measurement and statistical significance, effects of confounding events and environments, and additional analytical concerns.

In the area of measurement and statistical significance, TCRP 95 finds that the most common approach to evaluating change is the before-and-after comparison of traffic counts, transit passengers, or travel surveys. Before-and-after results are sensitive to the conditions at the two specific time points chosen, and often neglect to collect information on or consider the fluctuations around the average conditions. Olsson et al. [57] confirm these limitations, finding that the selection of reference time points can change not only the magnitude, but also the direction of the measured effects.

The second issue highlighted by TCRP 95 is confounding effects and environments. These include changes such as competing or complementary transport projects, employment and economic changes, and underlying socio-demographic trends. The report's recommendation that such factors be documented makes it clear that in many cases, those factors are not even acknowledged, much less analysed in a statistically rigorous manner.

Within the topic of "additional analytical concerns", TCRP 95 highlights the reliance on cross-sectional data, both in modelling and in comparative analysis. Cross-sectional data is limiting because it implicitly assumes that the behaviour of travellers remains consistent through time, and because self-selection and self-sorting issues (such as travellers who prefer to use transit choosing to live closer to rail stations) make it difficult to distinguish between correlation and causality.

1.1.3 Current State of the Field

In summary, the profession of transport planning is currently in a position where we recognise that travel demand forecasting accuracy is often a problem, but with the exception of a few limited case studies, we do not understand the causes of that problem. Furthermore, we lack the data to understand the problem. The systematic ex-post evaluation of transport projects has been identified as a means both to understand the problem and build a body of reference cases for comparison. There is increasing interest in conducting such studies, but those studies are hampered both by the burden of data collection and by methodological challenges in the analysis of longitudinal data in a system where there are always confounding effects.

1.1.4 The Emerging Opportunities of Big Data

In parallel to this increased interest in the ex-post evaluation of transport projects, a generation of data sources is coming on-line that provide another view of the transport system. These data include sources such as Global Positioning System (GPS) traces, mobile phone location data, transit smart card transactions, Automated Traffic Recorder (ATR) data, Automated Vehicle Location (AVL) data, Automated Passenger Counter (APC) data, new administrative sources of land-use characteristics, and a range of others. These "Big Data" provide new opportunities to observe the transport system.

Shuldiner and Shuldiner [58] examine the relationship between data and models, and the evolution of both over the history of travel demand modelling. They conclude:

> "From its very outset, travel demand modelling has been a data-driven activity; it has also been a data-restricted activity. But in recent years the nature of this relationship has undergone a profound change. Contrast, for example, the introduction of the home interview travel survey in the 1940s with the current use of real time location data. In the first instance, a data collection technique was developed and applied to meet a specific need—forecasting future travel demands. These data, collected by public agencies to meet a public purpose, then formed the basis for most institutional transportation planning and academic research for the next 50 years. Today individual trip-making data, in contrast, are collected primarily by private firms for their own corporate purposes, be it to provide a service to travellers or to sell advertising to third parties. Public agencies are engaged in similar, albeit less intrusive, activities through automated tolling and travel time monitoring programs. Technology—not theory or public purpose—is the driving

> force and has placed the individual squarely and firmly at the nexus of travel information, traffic management, and transportation planning. Technology per se is indifferent to the public interest; what really matters is the use to which technology is put.
>
> Transportation agencies and academic researchers now have available an unprecedented wealth of both activity and travel choice data with which to evaluate investment decisions and to devise models of travel behaviour. The challenge now is how to make the most effective use of these data."

A key distinction Shuldiner and Shuldiner make is that Big Data are often designed for another purpose, and their use in transport is secondary. Therefore, they bring about new challenges, both in terms of the methods for how to use these data, and governance issues related to their use for public versus private interests. With respect to governance issues, the goals of this thesis fall squarely in the realm of serving the public interest, while protecting the privacy of the data subjects. All software is being made open source, and the scientific results will be published, but no personally-identifiable information will be made available.

Regarding the methods for analysing these emerging Big Data sources, a good deal of progress has been made in recent years.

National Cooperative Highway Research Program (NCHRP) Report 775 [59, 60] examines the use of GPS data to understand travel behaviour. GPS traces of taxis are being used for traffic modelling [61, 62, 63, 64]. In addition to GPS, mobile phone data has been used to observe travel patterns [65, 66, 67, 68, 69, 70], and NCHRP Project 08-95 is further pursuing the problem [71]. Highway analysis has further focused on the use of Bluetooth data [72], electronic toll transponder data [73], Intelligent Transportation Systems (ITS) data [74, 72], and inductive loop detector data [75, 76]. Transit data analysis has focused on three core sources: AVL data [77, 78, 79, 80], APC data [81, 82, 83, 84], and transit farecard data [85, 86, 87, 88].

In addition to their independent use, GPS and other emerging technologies are also being used to enhance traditional travel survey methods, such as with the GPS-based Cleveland household travel survey [89]. Zmud et al. [90] compile a wide range of applications focused on the intersection of survey methods with Big Data from the proceedings of the 9th International Conference of Transport Survey Methods. There are more applications, but these examples give a flavour of the type of analysis that is being conducted.

In spite of these important achievements, there remain areas where further work would enhance the value of these data.

First is the integration of multiple data sets as a means for understanding and correcting for the biases inherent in any one data source. The studies cited above dominantly focus on a single data type, and because the sampling mechanism is often not designed with travel demand analysis in mind, the biases inherent in the data are often unknown. Smith [91] examines these issues in the context of the maturation of survey research several decades prior, and argues in favour of hybrid approaches that combine or cross-validate data sources in an effort to overcome the limitations of each. In a way, travel demand models have always done this, for example by estimating models from travel survey data, then validating them against traffic counts, so it is natural for travel demand modellers to continue in this direction.

The second area where the data offer largely untapped potential is in their use as a longitudinal data source. Often these studies are focused on developing the methods for analysing the data and using them to observe the current state of the system. Wang et al. [69] offer an example in their study which processes a month's worth of mobile phone data, a common approach in most of the studies reviewed. This is perfectly suitable for the stated goals, particularly as more data would often increase the cost or computational burden of the study. However, an important feature of these emerging data sources, in contrast to conducting custom-designed data collection efforts, is that they are often passively collected and provide a continuous stream of information.

Because of their continuously collected nature, these emerging Big Data sources provide a unique opportunity to study the effects of transport project, ex-post. By providing observations at more than one or two points in time, the trends in demand can be more closely related to other trends that may contribute to the differences observed. The key to taking advantage of these data is to develop a system where the incremental burden of adding more data through time is small.

1.2 Research Questions and Overall Approach

The overall approach taken to conducting this research aligns with two overarching research questions:

1. How can continuously collected data be leveraged to develop a data fusion tool suitable for monitoring travel demand trends?
2. How can the outputs of that tool be used to gain insight into the drivers of travel demand trends and to measure the transport project impacts?

The first question is focused on the methods and software needed to process the data and report meaningful performance measures. The output of this work is a prototype that can be run by a transport planner. It accepts new data as it becomes available, and can be run to provide reports and summaries for the available time periods.

The second question is addressed by applying the tool to study real-world transport change. A core component of answering this question is the application of statistical methods to estimate the treatment effect of the transport intervention, separate from other confounding factors.

The City and County of San Francisco serve as the context in which to answer both questions. The analysis focuses specifically on transit demand. Initial work has established the methods necessary for incorporating detailed highway data, but that work is not sufficiently complete for inclusion in this thesis. The time span covered varies according to the availability of data, going as far back as 2001, but focused in more detail on the 2009-2013 period where the most detailed data are available.

San Francisco was selected as as the test site for this work because the San Francisco County Transportation Authority funded the initial research. In addition, it is a good place to explore these issues for several reasons. First, it is a large city with a rich transit system, allowing transit issues to be analysed in a location where they are a major component of the travel market. Second, it is an economically vibrant city and a centre for technology development, which presents an interest set of topics to explore. Third, the author previously spent several years developing travel models in the region, which made it easier to obtain the relevant data.

The transit analysis focuses on the two largest transit operators serving San Francisco County: the San Francisco Municipal Railway (MUNI) and the Bay Area Rapid Transit (BART). San Francisco is also served by commuter buses coming from the East Bay, North Bay and Peninsula, and by the Caltrain commuter rail which runs south to San Jose. The MUNI analysis focuses more specifically on the MUNI bus system than on the MUNI light rail system, because detailed continuously collected data are available for the buses, but not for the light rail vehicles.

The thesis specifically explores four related aspects of the broader problem, as described in Section 1.6, but first, the unifying aspects of the approach are examined in further detail in the remainder of this introductory chapter. Section 1.3 takes inventory of the data sources available to this study. Section 1.4 describes the overall system design, and Section 1.5 considers the methods available to analyse the data output from the tool.

The review of existing literature is integrated into each chapter of the thesis, allowing it to be presented closer to where it is relevant within the thesis. Literature searches relied primarily on TRID (https://trid.trb.org/), a database that integrates records from TRB's Transportation Research Information Services and OECD's International Transport Research Documentation database. This database is preferred because it is focused specifically on transportation, it is international in scope, and it contains records for "grey literature", such as government reports, in addition to indexed academic journals. Searches were specific to the topics discussed in each chapter, and the results were reviewed for relevance before inclusion in the thesis.

These were supplemented by additional materials identified through pre-existing knowledge and personal connections. Where published review papers were available, the thesis relies first on the published reviews, and supplements them with materials published after the review paper was completed.

1.3 Available Data

The core of this study is determining how to take advantage of longitudinal data sources to understand the transport system. Relevant data sources available to this study are described here, segmented by category.

1.3.1 Transit Data

The transit data include schedule, location, demand, and farecard data.

General Transit Feed Specification (GTFS). GTFS [92] is a data specification that allows transit operators to publish their schedules in a standard format. It is commonly used for mapping and route-finding applications, and the standardisation allows application developers to write one set of code that works for many agencies. It only covers the current schedule, but when a new version is published, the old tends to be archived [93], so the differences can be used to systematically identify transit service changes.

Automated Vehicle Location (AVL) and Automated Passenger Counter (APC) data. About 25% of the San Francisco Municipal Railway (MUNI) bus fleet is equipped with AVL/APC technology. The AVL records the location and timestamp of the vehicle arriving at and departing from each stop. The APC records the number of passengers boarding and alighting at each stop. The buses with AVL/APC equipment are randomly assigned to drivers and routes at the depot each day such that over a number of days all routes are observed. These data are supplemented with manual counts of rail ridership, because rail vehicles are not equipped.

BART monthly entry and exit matrices. The Bay Area Rapid Transit (BART) system is a heavy rail system serving the San Francisco Bay Area. It is a major carrier of commuters into San Francisco, particularly because it provides a transit alternative to the heavily congested Bay Bridge. It is a closed system using distance-based fares, so passengers must have their ticket read both upon entering and exiting the system. Using this information, BART publishes monthly matrices showing the number of trips by entry station and exit station.

Transit smart card. In 2010, the Bay Area introduced the Clipper Card, a smart card system that integrates payment across the major transit operators in the region. About 30% of patrons currently use the cards. The system records a timestamp, location, and fare type each time the card is used upon entering a vehicle, or in the case of BART, entering or exiting. They are valuable because they would provide information on transfers, allowing the analysis to distinguish between unlinked and linked transit trips.

1.3.2 Highway Data

The available highway data come from three primary sources, described below.

California Performance Monitoring System (PeMS). The California Department of Transportation (Caltrans) operates an extensive network of traffic detectors on state highways, primarily the freeways in San Francisco. Data from these detectors showing the volume, speed and density of traffic by lane is published online in the PeMS [94]. These data are archived, often going back as far as 2002. They provide a great deal of information on freeway conditions.

Probe Vehicles. A long time series of taxi GPS traces were obtained from the Cabspotting project [95]. Cabspotting was a "data art" project that created visualisations of taxi movements in San Francisco for an exhibit at the Exploratorium science museum. The primary use of these data is to be used as probe vehicles providing observations of traffic speeds on city streets where PeMS is unavailable.

Local Traffic Counts. Given that PeMS is not available on all local streets, it is desirable to have traffic counts on a broader range of locations. To do this, counts are available from CountDracula [96], a database that integrates counts from a wide variety of sources in San Francisco.

TTI Urban Mobility Scorecard. The Texas A&M Transportation Institute (TTI) publishes an annual mobility scorecard that reports reports a number of auto congestion metrics for urban areas in the US, with data going back to 1982 [97]. The metrics include total congestion delay, wasted fuel, the monetary cost of congestion, and a travel time index, which is the ratio of travel time in the peak period to travel time at free flow conditions.

TomTom Traffic Index. TomTom is a maker of car navigation devices, but also uses data collected from those devices to monitor speeds. They report a congestion index for major metropolitan areas, which is a measure of the extra travel time attributable to congestion [98]. San Francisco data are available annually from 2008 through 2015.

1.3.3 Data Sources for Demand Drivers

In addition to observations of the transport system itself, it is important to understand the factors that drive demand for that system. Throughout this thesis, the term "drivers" is used to indicate factors that contribute to a change in demand. This terminology is selected to avoid indicating a formal causal relationship, although in many situations, there are sound theoretical reasons to believe that the relationship should be causal. In the event that it instead refers to a person driving a vehicle, it will be specified at that point in the text and through the context. Key data sources for observing those drivers are listed here.

LEHD Origin-Destination Employment Statistics (LODES). Traditionally, base year employment data used by travel models is derived from either state unemployment insurance records or commercial market research listings [99]. These sources often require a good deal of cleaning and error checking, specifically to resolve headquarters issues [100] (such as all the employees of

McDonald's being listed at a corporate headquarters rather than at individual stores). Therefore, it is common to only update the employment database in five year increments. More recently, the US Census Bureau began producing the Longitudinal Employer-Household Dynamics (LEHD), which combines these same unemployment insurance records with other data sources to report workforce indicators quarterly at the county level. Building upon the LEHD, it began producing the LEHD Origin-Destination Employment Statistics (LODES), which reports employed residents, employment and worker flows down to a Census block level [101]. LODES is available for 49 of 50 states, and is updated annually. This provides a new opportunity to monitor the longitudinal changes in the level and spatial distribution of employment.

2010 Census. While it is not longitudinal, the 2010 Census provides a reliable observation of population, households and housing units that serves as the starting point from which to measure changes.

Planning Department Building Completion Database. The City of San Francisco has provided a database of residential building completions that includes the address, opening date, and number of units in any newly completed residential building. It also includes information on any housing units removed, allowing the net change can be monitored.

American Community Survey (ACS). The US Census Bureau conducts an annual survey of 1% of households to collect information such as household income, household composition, number of vehicles, and so forth [102]. These data provide some ability to monitor trends in these measures, although at a spatially aggregate level.

Fuel Price. The US Energy Information Administration (EIA) regularly updates and publishes the average cost per gallon of gasoline and diesel fuel sold by location [103].

Mileage Rates. Recognizing that the cost to operate a vehicle varies not only with fuel cost, but also with the vehicle efficiency, some measure of those trade-offs is valuable. The Internal Revenue Service (IRS) provides this in their rates at which mileage driven can be reimbursed [104]. Separate rates are available for travel made for business purposes, versus travel made for medical visits or to move to a new city, all of which are tax deductible. The business rate reflects the full cost of owning and operating a vehicle, while the medical and moving rate represents the IRS' assessment of the average cost to operate a vehicle, exclusive of ownership costs. These values are generally updated annually, with previous values archived.

Toll Schedules. Toll schedules for the bridges in the area were obtained from the Bay Area Toll Authority (BATA) [105].

Parking Cost. Parking cost can be an important determinant of mode choice, and also can vary significantly with changes in employment. The most reliable way to monitor changes in parking costs is probably to walk the streets and physically record the posted rates. Lacking the resources for such a labour-intensive approach, this project instead relies on real-time data feeds of the availability and cost of city-owned parking spaces. This information is available through the SFPark Application Program Interface (API) [106]. In addition, SFPark has recently conducted a citywide parking census, which provides an observation of base conditions.

Consumer Price Index (CPI). Given that prices are used, the Consumer Price Index (CPI) is used to adjust all prices for inflation [107].

1.3.4 Other Related Data

While they are not directly incorporated into the data fusion tool described in Chapter 3, there are several additional data resources available to the project that can provide useful reference points for additional analysis or validation.

California Household Travel Survey (CHTS). The California Household Travel Survey (CHTS) [108] is a large-scale travel survey collected in California from 2010-2012, with oversamples in the urban areas. In addition to a travel diary, it includes GPS data collection for a subset of households. The combination provides a richness of information on individual travel behaviour that is beyond what is available from the other sources, so it may be used as a tool to enrich or expand other sources. It is not a longitudinal data sources, though, so it does not fit within the core functionality of the data fusion tool.

Mobile Phone Derived Trip Tables. Trip tables derived from mobile phone location data were obtained from two sources. Unfortunately, these data are not available longitudinally, so can only be used as a check of or starting pivot point with respect to the other data.

SF-CHAMP Travel Model. San Francisco Chained Activity Modeling Process (SF-CHAMP) is an activity-based travel model for San Francisco and the surrounding Bay Area. It operates using a microsimulation (agent-based) approach to simulate the movement of individual travellers throughout the region, and includes special functionality to consider pricing, bicycling and transit crowding [109, 110, 111]. The model inputs and outputs provide a useful picture of the transport system demand and supply.

DTA Anyway. Complementing SF-CHAMP is a citywide Dynamic Traffic Assignment (DTA) model known as DTA Anyway [112]. A core component of DTA Anyway is a set of tools to integrate network information from several sources, with the inclusion of observed timing plans for all 1,100 traffic signals in the city a major feat. These data are also available to the project, but they are not explicitly tracked longitudinally.

1.4 Data Fusion and Performance Monitoring

The first phase of this work seeks to answer the first research question: "How can continuously collected data be leveraged to develop a data fusion tool suitable for monitoring travel demand trends?" This question is answered by developing a prototype data fusion tool for the City of San Francisco, using that prototype to identify lessons learned and key concepts for the future generalization of such tools and transfer to other cities. This section describes the basic design of that system.

1.4.1 Key Design Goals

The data fusion tool seeks to fulfil five key design goals:

Usable. The target user group for this tool is professional transport planners or modellers, and the tool seeks to be usable by that group. This particular user group is well-placed to take advantage of the technology because they are a group that decision-makers currently turn to for analysis of complex planning issues, so they bring an existing level of credibility within the planning process. In addition, this group tends to bring an existing degree of comfort working with data and models, making them a more sophisticated user group than the general public. The system design seeks to be highly usable to this group, saving them effort from what would otherwise be a burdensome data assembly exercise, and giving them incentive to adopt the tool. One of the most important ways in which this is achieved is by encapsulating what would often be manual data processing interventions into code or control files. In this way, it follows the model of other data tools [113], to avoid a common source of error, document the process, and make the results reproducible.

Representative. An important limitation of Big Data is that because they are often used for a different purpose than they were originally collected for, there is a risk that they are biased in a way for which a well-designed survey would be able to adjust [91]. This tool seeks to mitigate such biases wherever possible, and explicitly acknowledge them otherwise. It does this through the comparison of and expansion to overlapping data sources where they are available. For example, transit AVL and APC data are available for a sample of approximately 25% of city buses, but schedule data (GTFS) is available for all buses. The AVL/APC data are expanded by route, direction, stop, month, day-of-week and time-of-day to match the total number of bus trips as observed in the GTFS data. This is similar to the process by which an onboard transit survey would be expanded to match ridership counts by route, or a household travel survey would be expanded to match census control totals by demographic group.

Consistent. Because the explicit goal of this tool is to measure change, the consistency of the measurements over time is of particular importance. Therefore, care is taken to ensure that the measurements are consistent even if data are missing for certain periods, or the penetration rate of certain technologies, such as the adoption of transit smart cards, changes over time.

Multi-level. Keeping in mind the target users, this system is designed to serve as a multi-level analysis tool. This means that it automatically reports a set of standard performance reports, but allows the user to run custom queries on the full data set. In this way, it both recognises the tension between being complete and overwhelming the user with details, and acknowledges that not every possible use will be anticipated a priori.

Transparent and transferable. The prototype is developed for a single city, the ultimate goal is that the tool can be modified and re-deployed elsewhere. This is facilitated by the publication of the software via an open-source licence. While there will be some cost to transferring the software and customizing it to work with the data sets available in other cities, that effort is left as a future development exercise, with this research serving to demonstrate the feasibility of the process in the first instance.

1.4.2 Data Flow

Figure 1.3 shows the basic system design with respect to the flow of data.

Starting from the left, there are two mechanisms for inputting data to the system. The first is “dump and transfer”, which simply means that the data are exported in a batch form and transferred into the system as a file. The transfer can be via email, disk, download, or a variety of other formats, but generally requires manual intervention to update. In many cases this approach is required when the data

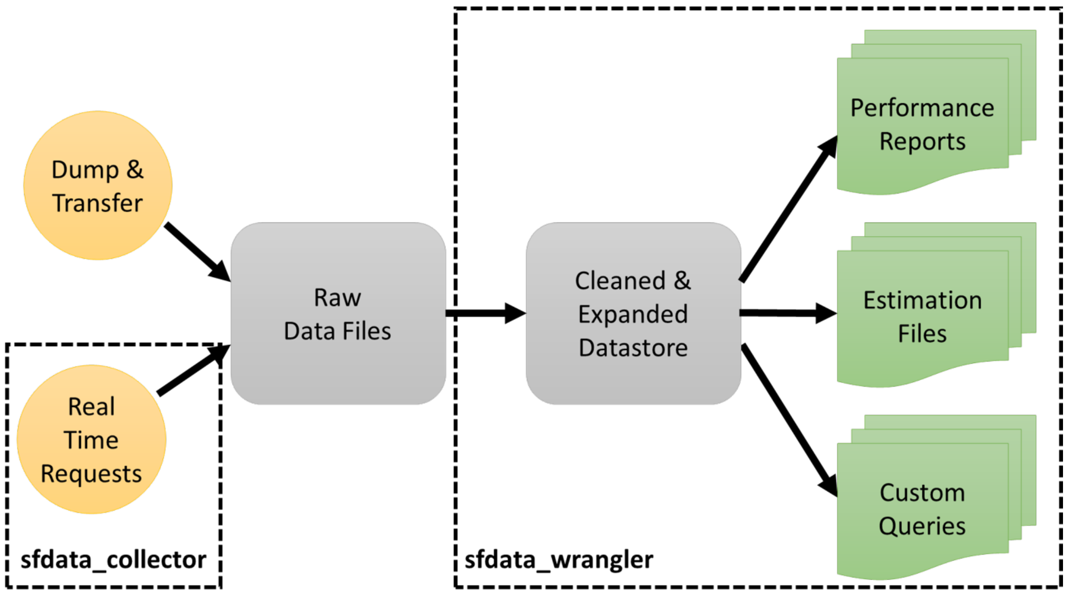


Figure 1.3: Data fusion tool design

are stored on a private server or in a proprietary format, but as long as future data remain in the same format, the effort required to update the system remains small. The second way to input data is via real time requests. These are handled through the `sfdata_collector` software [114] that was developed for this project. In cases where real-time data are provided via an API, `sfdata_collector` requests those data at a specified time interval and stores them to a database. This provides a means by which to archive those data, which would otherwise not persist, allowing the trends to be analysed. `sfdata_collector` is currently operational for the SFPark parking information and pricing system [106], but can be expanded to other real-time data sources using the same structure.

The data are stored on disk in their raw format, which varies by the type of data. They are converted to a processed datastore by the `sfdata_wrangler` package [115]. This involves a series of steps which include converting the data to a common format, identifying and handling problem records, adding derived data fields, expanding and weighting less complete data sources to be consistent with more complete sources, and adding aggregated tables (while retaining the disaggregate data). Chapter 3 gives more details of what is done for the initial data sets. The processed data are stored on disk using the Hierarchical Data Format 5

(HDF5) [116], which allows the fast storage of and random access to very large data files.

`sfdata_wrangler` performs three additional functions for data output, as shown on the right-hand side of Figure 1.3. First, it produces a set of performance reports at various levels of aggregation. Examples include monthly summaries of transit system ridership and level-of-service, and route-level demand profiles. Complementary software [117] can be used to map these outputs. Second, the software generates estimation files, which are in a structured format allowing statistical models to be estimated. Third, the processed data can be queried directly for custom tabulations and extractions. This is handled through the `pandas` package [118] which provides a range of data querying and analysis functionality that integrates cleanly with the HDF5 datastore.

1.5 Analysis of Contributing Factors

The second phase of work seeks to answer the question: "How can the outputs of that tool be used to gain insight into the drivers of travel demand trends and to measure the transport project impacts?" In doing so, it seeks to measure the treatment effect of transport projects, separate from other confounding factors, such as changes in employment levels or changes to competing transport service.

1.5.1 Confounding Factors in Travel Demand

The most common type of ex-post evaluation in travel demand is a before-and-after study, which simply compares measures for a time period before the project to a time period after the project [119]. In traffic safety, where ex-post studies are commonly used to evaluate safety improvements, this design is referred to as a naive before-and-after study, and it is recognised to have difficulty distinguishing between the treatment effect, and other effects, including changes to the exposure level, underlying trends and random noise [120]. In travel behaviour, such factors are sometimes discussed as considerations, but are rarely quantified [39]. The result is that such studies are at risk of errors in the magnitude or even the direction of the estimated project effect [119, 57].

Observing these challenges, this research takes the position that a robust ex-post analysis in transport should consider the following factors:

Treatment effect. The change attributable to the project itself.

Exposure effect. Changes attributable to an increase or decrease in the number of travellers exposed to the project. This would manifest as a change in the level of or spatial distribution of population and employment in the surrounding area.

Trend effect. Underlying trends that are not accounted for in observed factors. These might include demographic changes reflected in the observed variables, seasonality, or cohort effects.

Random effect. Noise in the data due to natural variation, special events, or other factors which are not directly measured.

Other observable effects. Non-project factors that can be observed and included directly in the analysis. These might include changes in fuel cost, parking cost, competition with cars and competition with other transit routes.

1.5.2 Methods for Controlling Confounding Factors

A number of methods have been used in past transport studies to control for confounding factors such as these. These methods can be grouped into four main categories: control groups, modelling the counter-factual, forms of regression, and methods to deal with unobserved confounders. For a general discussion of such methods for use in policy evaluation, see Coglianese [121], and for a discussion of their application to transport, see Haight [119].

The first approach for accounting for these factors is to use control groups. Tay et al. [122] provide an example of how this can be done in the context of a before-and-after study. They examine the change in neighbourhood crime after the introduction of rail stations, using the remainder of the city as a form of control group. In their study of the effect of three types of policy on transit use, de Grange et al. [123] employ a slightly different strategy of examining changes across 41 different cities, with the cities serving as a form of control. While these examples provide a form of control group, they are not true randomised control trials. In the context of health research, Ogilvie et al. [124] do identify a number of studies of interventions to promote walking that are randomised control trials, although it is clear from the review that such strict methodological criteria limit the types of policies that can be practically evaluated.

The second approach to separating the project treatment effect from other confounding factors is to model the counter-factual. The counter-factual is what would have happened had the project not been implemented. The EDR Group [125, 126] provides an example of this approach in their evaluation of Boston's Central Artery/Tunnel project. The Central Artery/Tunnel, also known as the Big Dig, was the most expensive infrastructure project in US history, costing $15 billion, which was triple the original estimate. However, EDR found that the mobility benefits were 15-20 % higher than originally forecast, and the urban redevelopment benefits were 10 times higher than estimated [127]. These estimates were derived by updating an existing travel model and economic model with observed data, and using them to model current conditions without the project. The prerequisite for such an approach is the existence of these models, as well as data with which to update them.

The third approach used to controlling confounding factors is the estimation of statistical regression models. A number of forms of such models have been used in transport studies with a time-element, generally falling into the categories of either time series models or longitudinal models. The distinctions lie in the number of entities observed, and the number of times each is observed. Traditionally, time series data referred to a single entity observed many times [128], such as Chen et al.'s [129] study of total transit system ridership. Longitudinal data has multiple entities observed multiple times. This can be Time Series Cross Section (TCSC) data where there are a small number of units observed many times, or panel data where there are a large number of units observed a small number of times each [130]. An example of TCSC data is Tang and Thakuriah's [131] model of the effects of traveller information on bus ridership by route in Chicago. Frazier and Kockelman [132] use spatial panel approaches to model land cover change in Texas. These models can be estimated with or without spatial correlations. The Frazier and Kockelman models are spatial, and Cheng et al. [133] demonstrate the use of spatial time series models and also provide an example of the application of time series models to data from multiple sensors. There can be overlap in the applicability of these methods depending on how large and small are defined with respect to both observational units and temporal units. Often, the distinctions are by field, for example with health scientists referring to longitudinal data and social scientists referring to data of the same structure as panel data [130]. Within transport, there does not appear to be a consistent body of work to identify a norm. The common theme to all of these statistical models is that they provide a means by which to estimate the effect of the project, other control variables which can be directly included in the model, and some form of time trend.

The final approach to dealing with confounders is a suite of methods designed to handle unobserved confounding factors. These are also statistical models, derived from the health and social science fields. Examples include propensity scores, such as Graham [134] and difference-in-difference methods, such as Li et al. [135]. The advantage of these methods is that they deal with unobserved confounders that can arise from the non-random assignment of treatment [121].

1.5.3 Selected Approach

For this study, data across multiple cities and true random control groups are not practical. It would be possible to treat other parts of the same city as a control, although this can be integrated into the statistical modelling approach discussed below. Therefore, the control group approach is not used explicitly here.

In the context of this research, modelling the counter-factual is both a practical and appealing approach. In the context of evaluating a transport project, either the model that was used to originally develop forecasts for the project, or a newer version should be available. In addition, the data fusion tool generates a continuous series of data that can be used as input to the model for the appropriate dates. The appeal of this approach is that it would take advantage of a full-scale transport model that is likely to be behaviourally and theoretically more sound than could be estimated directly from the longitudinal data. The disadvantage to this approach is that it is not independent, and potentially subject to the same limitations as the original forecasts.

The statistical modelling approach is appealing in this context because it provides an independent means to estimate the project impact directly from the longitudinal data. It offers an important step forward over a naive before-and-after study design, and provides a mechanism to separate the treatment effect from the exposure effect, trend effect, treatment effect, random effect, and other controllable factors.

The approaches to dealing with unobserved confounders may offer some advantage, because transport projects tend not to be randomly assigned (unless you think very poorly of the political process involved in selecting them). However, the focus of this research is on assembling the data such that as many of the critical factors as possible can be included as observed confounders. Therefore, these approaches to dealing with unobserved confounders are not prioritised within this study.

So we are left to consider two good, but different approaches: model the counter-factual using an existing transport model or estimate statistical models directly from the longitudinal data. In this project, we choose to pursue the latter because it 1) provides independent estimate and 2) it is able to utilise the full time-spectrum of data assembled, and not just the data for the year of the counter-factual. This provides both a stand-alone mechanism to estimate the project effects using data that is being automatically collected and processed through the data fusion tool. Time series models, specifically, are used here, although panel data models offer a promising avenue for future research.

1.6 Thesis Structure and Contribution

The structure of the thesis and the broad research contributions are described below.

1.6.1 Structure

The remainder of this thesis is structured around five chapters that explore specific aspects of the problem.

Chapter 2 examines, in detail, the biases and limitations of one particular data set, transaction data from the Bay Area's transit smart card system, Clipper Card. In contrast to much of the previous research that uses smart card data, the Clipper data is subject to strong privacy restrictions, less complete data fields, and a lower penetration rate. If Big Data methods are to be applied more broadly, it is important to consider how they may apply to imperfect data sources. This chapter explores that question, examining the remaining value of the Clipper data and the biases inherent in those data.

Chapter 3 describes the development and features of the core software tool used in this research. It focuses, in particular, on the process for merging transit AVL/APC data with GTFS data, and weighting the former to make it representative of the full system ridership. It discusses how the software tool may be useful for performance based planning.

Chapter 4 uses the outputs of the data fusion tool to estimate time series models of ridership on two transit systems: the MUNI bus system and the BART rapid transit system. These models account both for service changes and external factors expected to affect transit ridership.

In Chapter 5, the estimated models are used to explore and explain the divergent ridership trends of the two systems, with BART experiencing strong ridership growth while MUNI ridership stagnates.

Chapter 6 starts from these same time series models, and demonstrates how they can be used in five example applications. Three of the applications relate to the ex-post evaluation of transport changes. For these, the ridership effect measured with the time series models is compared to the ridership effect implied by a naive before-and-after comparison, and to the effect implied by the application of published elasticities. Two of the applications demonstrate how short term forecasts derived from the models may be useful in performance based planning.

The thesis finishes with conclusions, lessons learned and next steps. These relate both to the specific topics explored in the chapters, and the broader themes described in this introduction.

Supplemental material is included in several appendices. Appendix A identifies related works that are based on the content of this thesis, and describes the contributions made by co-authors of those works. Appendix B provides a full enumeration of the data summarised at the beginning of Chapter 4, and Appendix C shows the formulas that can be used to apply the time series models estimated in Chapter 4.

1.6.2 Research Contribution

This research addresses the two areas of need described at the end of Section 1.1.

It places a strong emphasis on examining multiple data sources in relation to each other, both for the purpose of identifying important biases, and with the goal of combining the data in a way to correct for those biases. For example, Chapter 2 identifies socio-economic biases in the available transit smart card data by comparing them to onboard transit survey data, and estimates a set of correction factors to mitigate those biases. Chapter 3 describes how the software component of this project links sampled transit vehicle location and passenger count data with a full enumeration of scheduled operations to develop a weighting scheme for the former.

In addition, this research is longitudinal in nature, focusing specifically on data available through time. The longitudinal nature of the tool allows for a more complete analysis of factors contributing to the changes in demand, which can be obscured in a basic before-and-after study approach. The value of this approach is demonstrated through the application of time series models in Chapter 5 and Chapter 6.

By compiling the data necessary to better understand travel demand changes, this research lays the groundwork for two important outcomes. First, it provides the data necessary for the evaluation of past travel demand forecasts. Second, it provides the tools and analytical framework necessary to build a library of empirical case studies of transport projects. Together, these outcomes will allow better informed decisions about future transport investments.

Chapter 2

Transit Smart Card Data Evaluation

A core data source identified for use in this research is transaction data from the Bay Area's transit smart card system, Clipper Card. Chapter 3 will describe the development of a data fusion tool aimed at measuring transit system performance, with Figure 3.1 showing how Clipper data fits into that tool. However, when the Clipper data were obtained, it was with several limitations not present in some of the earlier published work exploring the uses of transit smart card data. These limitations include a data obfuscation process to protect privacy, a limited penetration rate, and some key fields that are often missing. This opened up a different line of research aimed at understanding the remaining value of data with such restrictions. This is important because it helps define how broadly published methods can be applied when some agencies have tighter privacy controls or different technology, and how agencies can plan to maximise the value of data that may emanate from the systems they operate.

2.1 Introduction

In recent years, transaction data from transit smart card systems have been used in a range of transport planning applications. Pelletier et al. [87] provide an overview of a range of those applications, including strategic, tactical and operational-level studies. Since the time of that review, smart card data have continued to be used in a number of cities. In Singapore, Medina and Erath used smart card data to estimate workplace capacities for input to a MATSim model [136]. Chen et al. use

Singapore smart card data in detecting the dynamics of urban structure [137], and also to evaluate the day to day variability in mobility patterns [138]. Munizaga and Palma used Santiago data to estimate disaggregate transit origin-destination matrices [139]. Gordon et al. developed methods to infer linked transit journeys using Oyster card data in London in combination with vehicle location [88]. Wang et al. used smart card data from the Beijing Metro to evaluate the fare changes [140].

A commonality among these studies is that they tend to be based on systems where the data are relatively complete, both in terms of the information recorded and in the penetration rate among transit patrons. For example, in Singapore, the data cover 97% of all transit trips, with both a tap-in and tap-out required for adult users [136]. The Santiago system also achieves a penetration rate of 97%, with the database recording the exact time of the tap-in and either the station code or a bus vehicle ID [139]. The recorded bus ID is important because it allows the transaction to be matched to the location of the vehicle, giving the boarding location. Similarly, Oyster cards in London record the time of the transaction, and vehicle information for buses, allowing the location to be derived by matching to vehicle location data. Oyster is used by 90% of bus riders, and 80% of riders on the London Underground, the latter of which requires both tap-in and tap-out at station locations [88]. The Beijing data record both the entry and exit location and the line used, and are taken in the study as an enumeration of the full demand in the system [140].

Collectively, these studies demonstrate that smart card data can have value in understanding the spatial distribution of transit trips, trip-linking behaviour, and measuring the effect of system changes. It is natural to seek to apply these methods in other regions, but the data in other regions may be of variable quality. Smart card systems could have lower penetration rates, local rules could impose stronger privacy restrictions, or the method in which the cards are used or the data are recorded could impose further limitations. In such cases, can the data still be used in the same ways, are they a valid representation of travel on the transit system as a whole, and can any biases be corrected in a systematic way?

This chapter seeks to address these questions through an examination of the

Clipper Card system in the San Francisco Bay Area. Clipper Card transaction data provide a good opportunity to study these issues for several reasons. First, it covers eight separate transit operators in the Bay Area, operating six distinct transit modes, allowing for a comparison across different operators and payment policies within the same database. Second, it is subject to a series of data obfuscation steps to more strongly protect user privacy, as discussed by Ory [141]. Third, the penetration rate is much lower than the studies cited above, at about 45% of average weekday boardings.

The research contribution of this chapter is to evaluate the value of smart card data with these limitations, and validate key dimensions of the data against available external data sources. It proposes a method to mitigate the biases in smart card data by estimating a discrete choice model of smart card use and applying the reciprocal of the modelled probability of using a card as a correction factor. The chapter goes on to provide information and recommendations to agencies on how to maximise the value of their own data, with varying limitations.

The remainder of this chapter is structured as follows: Section 2.2 provides the context of the Bay Area transit system and the Clipper Card payment system. Section 2.3 describes the data used in this study, including the Clipper data and supporting data sets. Section 2.4 evaluates the data along several key dimensions. Section 2.5 presents the estimated choice models and the result of their application as correction factors. Finally, Section 2.6 provides conclusions and recommendations for other agencies seeking to use such data.

2.2 Context

The 9-county San Francisco Bay area covers approximately 7,000 square miles with a population of 7.5 million residents. It has three core cities of San Francisco, Oakland and San Jose, and features geographic constraints in the Bay itself as well as surrounding mountains. Figure 2.1 shows an overview of the transit system, which includes the Bay Area Rapid Transit (BART) system, the Caltrain commuter rail, light rail systems in San Francisco and San Jose, major bus systems in the core cities, and a number of smaller operators.

Travellers are faced with navigating this system that is both geographically dispersed and operated by separate agencies with their own policies and payment methods. The Clipper Card was introduced in 2010 to provide common fare media for transit trips in the Bay Area. Clipper is a contactless smart card upon which travellers can load either cash value or a transit passes. Users are required to tag-on as they enter the transit system, and for modes with a zone-based fare, they are required to tag-off as well. Clipper is currently accepted by all of the major and some of the minor transit operators in the Bay Area. As the regional planning agency, the Metropolitan Transportation Commission (MTC) coordinates the system among the transit operators and houses the resulting transaction data. Planning for the next generation Clipper system is currently underway.

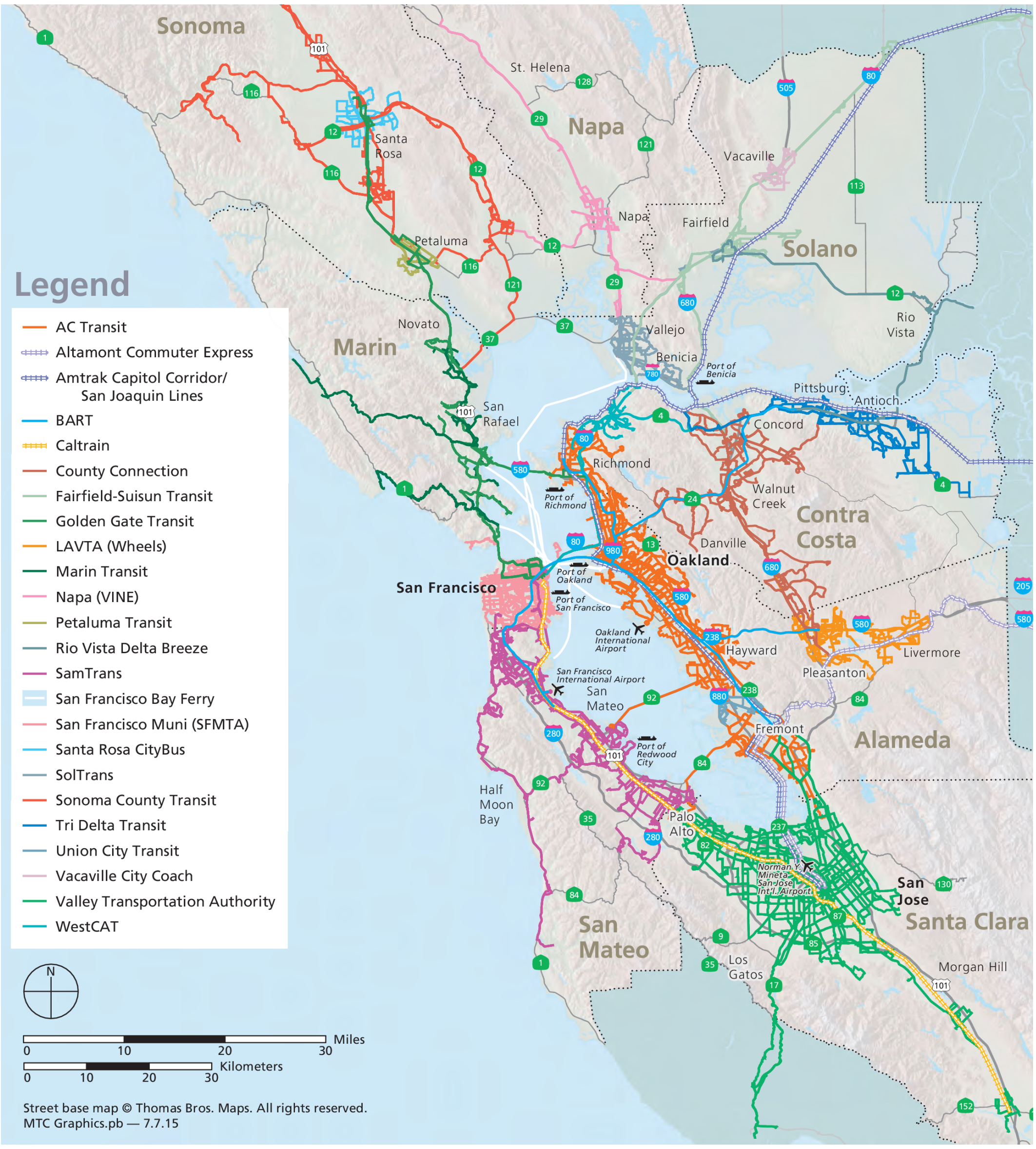


Figure 2.1: Bay Area transit operators [142]

2.3 Data

The primary data source used in this study is an anonymous database of Clipper Card transactions. In addition, several supporting data sources are used for the purpose of validation. All are described here.

2.3.1 Clipper Card Data

MTC has provided an anonymous database of Clipper Card transactions covering the period from March 2013 to September 2014. The data include one record each time a user tags onto the transit system, with the record including information on the card ID, the agency and payment type, the tag-on time and location, and the tag-off time and location where applicable. Table 2.1 describes the fields in the database.

The card technology is such that transactions made on a vehicle are not processed until after the data are downloaded when the vehicle is back at the station. Therefore, while the time of the transaction is known, the location of that vehicle at that time is unknown. The location of transactions made at stations, generally on rail, is known.

To protect users' Personally Identifiable Information (PII) and comply with California's strong privacy laws, MTC developed a multi-step data obfuscation process as follows [141]:

1. Starting from the universe of Clipper transactions, the Clipper Card ID is replaced with a random number that persists for one day. Therefore, within a day it is possible to identify the transactions made by the same card, but it is not possible to identify the transactions made by the same card on a different day.
2. Discard a random selection of 50% of card IDs for each day.
3. The transaction times are grouped into 10 minute increments.
4. The transaction dates are obfuscated. Each Sunday within each month is assigned a random number between one and ten. Three Sundays are randomly selected, retaining the random identifying integer, and the fourth (and fifth, if relevant) is discarded. This process is repeated for each day of the week. The

Table 2.1: Clipper data dictionary

Field name	Type	Example	Description	Notes
Year	smallint	2013	Transaction Year	
Month	smallint	10	Transaction Month (1 is January)	
DayOfWeekID	smallint	4	Transaction Day of Week Integer	A day is defined as 3 am to 3 am the following day
DayOfWeek	char	Wednesday	Transaction Day of Week Name	A day is defined as 3 am to 3 am the following day
RandomWeekID	smallint	6	Random Integer that Identifies a Unique Day	The Year, Month, DayOfWeek, and RandomWeekID fields uniquely identify a day
ClipperCardID	varbinary	D88268EA105â€¦	Anonymised Clipper card identifier	A random number representing a unique Clipper Card that persists for one circadian day (3 am to 3 am)
TripSequenceID	bigint	2	Circadian Day Trip Sequence	
AgencyID	int	1	Transit Agency Integer	
AgencyName	char	AC Transit	Transit Agency Name	
PaymentProductID	int	119	Payment Product Integer	
PaymentProductName	char	AC Transit Adult local (31 Day Rolling) pass	Payment Product Name	
FareAmount	money	0	Fare	Monthly pass holders have a zero fare for each transaction
TagOnTime_Time	time	17:35:00	Boarding Tag Time	Times are rounded down to the nearest ten minute interval
TagOnLocationId	int	2	Boarding Tag Location Integer	
TagOnLocationName	char	Transbay Terminal	Boarding Tag Location Name	
RouteID	int	300	Route Integer	
RouteName	char	F	Route Name	Not all bus operators transmit route names, e.g. all MUNI routes are recorded as 'SFM bus'
TagOffTime_Time	time	20:20:00	Alighting Tag Time	Times are rounded down to the nearest ten minute interval
TagOffLocationId	int	15	Alighting Tag Location Integer	For systems that require passengers to tag out of the system
TagOffLocationName	char	Millbrae (Caltrain)	Alighting Tag Location Name	For systems that require passengers to tag out of the system

result is that the data reveal that a transaction may have occurred on a Sunday in October, but not which Sunday.

5. The data are released in monthly chunks, with a lag in the release of new data.

There are two specific points of concern that these processes are designed to prevent. First, is avoiding the possibility of identifying individuals through their regular travel patterns. Second is a desire to avoid locating the card precisely in both time and space. The purpose of this data obfuscation is to make it impossible to identify an individual from the data even if it were combined with other sources and observations.

There is a trade-off between protecting privacy and having detailed data to analyse, and this process represents an initial attempt to balance these competing goals. MTC has asked for feedback on this balance, and this chapter presents evidence to provide a better understanding of what is lost on the data analysis side of that trade-off. By presenting this to the broader community, we invite others to comment based both on this evidence and their own experience.

2.3.2 Data Processing and Weighting

The Clipper Card data were processed to derive additional fields used in this research, including the mode, a transfer flag, and weights.

The transit mode was inferred based first on the agency name for agencies that operate a single mode. These include rapid transit for BART, commuter rail for Caltrain, Ferry for Golden Gate Ferry and SF Bay Ferry, and bus where not otherwise specified. For Valley Transportation Authority (VTA), transactions where the route name is LRV (Light Rail Vehicle) are coded as light rail, and others are coded as bus. MUNI includes bus, light rail and cable car. Any transaction occurring at a light rail station rather than on a vehicle was coded as light rail. Records with the route name equal to CC59, CC60 or CC61 were coded as cable car. Those with a route name of F, J, K, L, M and N were coded as light rail. The remaining MUNI records are coded as bus.

Transfers were identified in the system based on two tag-ons occurring less than 90 minutes apart. 90 minutes was selected as the duration free transfer on MUNI. When a transfer occurs, the transactions associated with that transfer (both before and after) are assigned the same linked trip ID such that the data can be summarised as linked trips in addition to boardings.

The data were weighted such that a month's worth of weighted records represent the transactions for an average weekday, Saturday or Sunday during that month. This involves scaling up by a factor of two to compensate for the 50% of card IDs that were discarded, and scaling down by the number of weekdays, Saturdays or Sundays in the monthly data set (15, three or three based on the number retained in the data obfuscation process). A linked trip weight is calculated that divides the base weight by the number of boardings in the linked trip. For example, a boarding on BART followed by a transfer to AC Transit would be counted as half a linked trip on each.

The calculation of modes, transfers and weights are imperfect, as discussed in the data evaluation section, but they provide a sufficient basis for conducting that evaluation. In particular, there is a risk that a changing penetration rate could be a source of error because an increase in penetration would look the same in the data as an increase in ridership. This risk is explored in Section 2.4.6.

2.3.3 Other Supporting Data

In addition to Clipper Card transactions, there are a number of other transit data sources in the Bay Area. Several are used in this study to validate key dimensions of the Clipper data. The Statistical Summary of Bay Area Transit Operators provides overall ridership numbers by operator and mode [142]. BART monthly ridership reports provide "entry/exit" matrices from the fare gates indicating total numbers of riders traveling between each station pair [143]. There is Automated Passenger Counter (APC) and Automated Vehicle Location (AVL) data for a sample of MUNI buses. These data have been expanded to the schedule information in GTFS (General Transit Feed Specification) format such that they are representative of total ridership, as described in Chapter 3.

Onboard transit surveys are available on Alameda County Transit (AC Transit) [144], SamTrans (San Mateo County Transit) [145], Golden Gate Ferry [146], Golden Gate Transit [147], San Francisco Bay Ferry [148], and Caltrain [149]. The onboard surveys are expanded to match boarding counts by route, direction and time-of-day. They are used throughout this chapter as a point of comparison against the Clipper data, with the implicit assumption that the expanded onboard survey data is a better representation of the true ridership than the Clipper data.

Given this assumption, it is worth acknowledging that the onboard survey data are subject to their own set of limitations. First, it is a sample, and subject to sampling error as a result (For these surveys, the goal was to collect a representative sample of 5% of riders aged 16 or older). Second, it is expanded to boarding counts, and is subject to any errors in those counts, which can be particularly challenging when the transit vehicles become very crowded. Even if the counts are perfect, there can be substantial day-to-day ridership variation, so there will be some variation depending on which day the counts were taken and which day the surveys were conducted. Third, there can be data quality and recording errors, particularly for manual transcription of paper-based surveys. Fourth, there can be non-response biases for certain categories of trips. For example, people making short trips may not have sufficient time to respond to a survey, and in a diverse, multi-lingual region, non-English speakers may be less likely to respond.

The methodology sections of the survey reports make it clear that the contractors understand these issues and have done their best to mitigate them. They employ a methodology that involves a very short questionnaire to collect origin, destination and contact information, followed by a telephone interview to collect more detailed information. This is aimed at minimizing non-response bias, and allows the telephone interview to be language-specific. The telephone interview features real-time mapping of the responses, which allows erroneous entries to be identified and corrected on-the-spot. In spite of these strategies, it is acknowledged that the onboard surveys are imperfect.

The validation is not comprehensive, but focuses on subsets of the Clipper data that overlap with the external sources. At the same time, if the Clipper data were to merely replicate existing sources, they would have limited additional value. So these data also provide a benchmark for assessing what the Clipper data can be used for over and above existing sources.

2.4 Data Evaluation

This section provides an evaluation of the Clipper Card data with respect to six dimensions: system characteristics, penetration rates, geographic distribution, temporal distribution, transfer rates and stability over time. This evaluation seeks to provide insight into three key questions:

- What new information do these data provide?
- Are they representative of the system as a whole?
- What can be done to make the data more useful?

Unless otherwise specified, all data are weighted transactions for an average weekday in March 2013. This month was selected because it is a month without major holidays, when schools are in session, and when MUNI AVL-APC data are also available.

2.4.1 System and Data Characteristics

The characteristics of the system itself, as well as details of how the data are coded are an important factor in determining what data actually get recorded and how they can be used. In the examples cited above [139, 88, 140], there is a clear distinction between the data collected on metro systems versus that collected on bus systems. The metros, like the London Underground, typically have fare barriers, and require the user to tap-in and tap-out, generating data with a known location at both ends of the trip. This structure generates the highest quality of data. By contrast, the buses require only a tap-in, and the location of that tap-in is not known by default but must be derived by matching that record to the location of the bus as recorded by AVL. That matching requires that the route name and the vehicle ID be recorded.

Table 2.2 lists the agencies and modes covered by the Clipper Card data, with selected characteristics of each. For the local bus systems (AC Transit, SamTrans, MUNI and VTA), as well as the MUNI cable cars, tag-ons happen on the vehicles, and the location is not recorded. It is worth noting that MUNI allows boardings through the rear door of the bus. Card readers are provided at the rear of the vehicle, although on a very crowded bus reaching them can be problematic.

Table 2.2: Transit system characteristics

Agency Name	Mode	Tag-On Required	Tag-Off Required	Tag Location	Fare Enforcement	Percent with Route ID Coded	Vehicle ID Provided
AC Transit	Bus	Yes	No	Vehicle	At Boarding	59%	No
BART	Rapid Transit	Yes	Yes	Station	Barriers	0%	No
Caltrain	Commuter Rail	Yes	Yes	Station	Proof-of-Payment	0%	No
Golden Gate Transit	Bus	Yes	Yes	Zone	At Boarding	100%	No
	Ferry	Yes	No	Station	At Boarding	100%	No
MUNI	Bus	Yes	No	Vehicle	Proof-of-Payment	6%	No
	Cable Car	Yes	No	Vehicle	At Boarding	100%	No
	Light Rail	Yes	No	Vehicle or Station	Partial Barriers	4%	No
SamTrans	Bus	Yes	No	Vehicle	At Boarding	3%	No
VTA	Bus	Yes	No	Vehicle	At Boarding	86%	No
	Light Rail	Yes	No	Station	Proof-of-Payment	0%	No
SF Bay Ferry	Ferry	Yes	Yes	Station	At Boarding	0%	No

The two light rail systems are both operated on a proof-of-payment system. On VTA light rail in San Jose, the card readers are on the station platforms, and users tap-in before boarding. Whereas most systems require a tap-in for a transfer, it is not required for VTA as long as the transfer time window has not expired. The MUNI light rail operates at the street level through much of San Francisco, and as a subway in the downtown area. Riders boarding at the street level tag-on to a reader on the vehicle. Riders boarding at the subway stations must pass through fare gates and tag-on at the station.

BART requires tap-ins and tap-outs at the stations with fare barriers, providing the most complete data set. The ferries are similar. While the Golden Gate Ferry does not require a tap-out, there is a single destination so it is unnecessary. Caltrain also requires a tap-out, but it is a proof-of-payment system. Golden Gate Transit buses are similar in that they require a tap-out to accommodate their zone-based fare structure, but because the readers are on the buses, the transaction location is only recorded at the level of a fare zone, not the specific location of the bus.

It is also important to note that the vehicle ID is not provided, and the route ID is left uncoded for many transactions. This, plus the fact that the transaction times and dates are deliberately muddied for privacy reasons, prevents the vehicle-based tag-ons from being matched to AVL data to derive their location. The dilemma is that the boarding location would be valuable information in terms of understanding the state of the system, but deriving that information would require locating a card precisely in both time and space. It is not an issue for the station-based data—the location is fixed so the time can be imprecise—but having a 10-minute window for an unspecific date makes it difficult to figure out where a vehicle is.

2.4.2 Penetration Rates

The penetration rate refers to the percent of transit users that pay using a smart card. While meaningful conclusions can be drawn from a relatively small sample of data, lower penetration rates mean that there is greater opportunity to magnify any biases in who uses clipper cards. That is, for systems like Santiago and Singapore with a 97% penetration rate, there is a limit to how far the results can be skewed.

Table 2.3 presents the weighted average weekday Clipper transactions for each Bay Area system in comparison to the total average weekday ridership [142]. The penetration rate is calculated as the ratio of the two, and varies from 8% on MUNI Cable Cars (with many tourists paying cash) to 91% on the Golden Gate Ferry. Recent onboard surveys on some of the systems have asked about Clipper usage, and the surveyed values are reported in the right-most column. Whereas the penetration rate calculated by these two methods compares reasonably well for most of the cases, the reason for the large difference on AC Transit is not clear.

Beyond the penetration rates themselves, it is important to understand what, if any, biases exist in who uses Clipper. Evidence in this regard is provided by the onboard surveys. The surveys show that Clipper Card use is correlated with employment, income, Hispanic identification and number of transfers, with employed travellers, higher-income travellers, non-Hispanics and trips with transfers more

Table 2.3: Estimated Clipper Card penetration rates

Agency Name	Mode	Weighted Clipper Transactions	Average Weekday Ridership	Penetration Rate	Penetration Rate from Onboard Survey
AC Transit	Bus	57,732	173,169	33%	58%
BART	Rapid Transit	214,724	420,396	51%	
Caltrain	Commuter Rail	8,487	49,031	17%	
Golden Gate Transit	Bus	9,053	11,986	76%	78%
	Ferry	6,775	7,465	91%	81%
MUNI	Bus	273,311	511,733	53%	
	Cable Car	1,678	20,523	8%	
	Light Rail	65,926	164,488	40%	
SamTrans	Bus	15,709	40,970	38%	45%
VTA	Bus	32,421	106,160	31%	
	Light Rail	5,824	33,730	17%	
SF Bay Ferry	Ferry	1,065	4,849	22%	36%
Total		692,703	1,544,500	45%	

likely to use Clipper [145, 146, 148].

The income correlation is pronounced. On AC Transit, Clipper usage ranges from 34% for household incomes less than $10,000 up to 71% for household incomes over $75,000 [144]. The bias is pronounced even with a reasonably high penetration rate. On Golden Gate Transit buses, whose 78% penetration rate is comparable to the Oyster Card usage on the London Underground, 33% of riders with income less than $10,000 use Clipper compared to 91% of riders with an come of $75,000 or more [147].

Unfortunately, it is not practical to directly weight the Clipper data to offset these biases, because doing so would require knowing the socio-economic characteristics of the card holders, and doing that would likely raise privacy concerns. Instead, we explore whether the Clipper data are representative of the spatial and temporal distributions of the full population of transit riders, with the idea that accounting for any biases in these dimensions could reduce any socio-economic biases. Results are presented in the following sections.

2.4.3 Geographic Distribution

To validate the geographic distribution of Clipper transactions, it would be best to compare to the geographic distribution of all trips from an independent data source.

While such a data source is not available for the Bay Area as a whole, it does exist for BART, which publishes a monthly ridership report showing the total entries and exits between each station pair based on fare gate data [143].

Figure 2.2 shows these fare gate totals plotted against the Clipper Card transactions between each station pair for an average weekday in March 2013. The Clipper values are lower because the penetration rate is only about 50%, but the correlation is strong. A line can be fit through the points with an R^2 of 0.95. The two largest outliers, located between 1,000 and 1,500 on the x-axis, are both directions between San Francisco Airport and Powell Street in downtown San Francisco.

Despite this good overall fit, inspecting the results at a district level reveals that the penetration rate varies in specific ways, with 63% penetration for trips within the core area of San Francisco and Oakland, 54% penetration for trips from elsewhere to that core, and 33% for trips outside that core. This trend is likely related to higher-income commuters coming travelling to those areas.

2.4.4 Temporal Distribution

The temporal distribution of the data were assessed for MUNI buses, because AVL/APC data were available for this system from which a detailed time-of-day profile could be derived. These data were reported using the data fusion tool developed in Chapter 3. This was done for average weekday conditions in March 2013.

Figure 2.3 shows the temporal distribution in 10-minute increments from these two data sources. The curves follow the same overall profile, with the primary difference that the Clipper data shows a much higher AM peak, with less mid-day travel. Again, this may be related to a higher rate of Clipper usage among higher-income commuters.

2.4.5 Transfer Rates

An advantages of transit smart card data is the ability to observe a full day's travel for a particular card, allowing transfers to be inferred. This is in contrast to other passive data collection systems, such as APCs, where it is not possible to identify

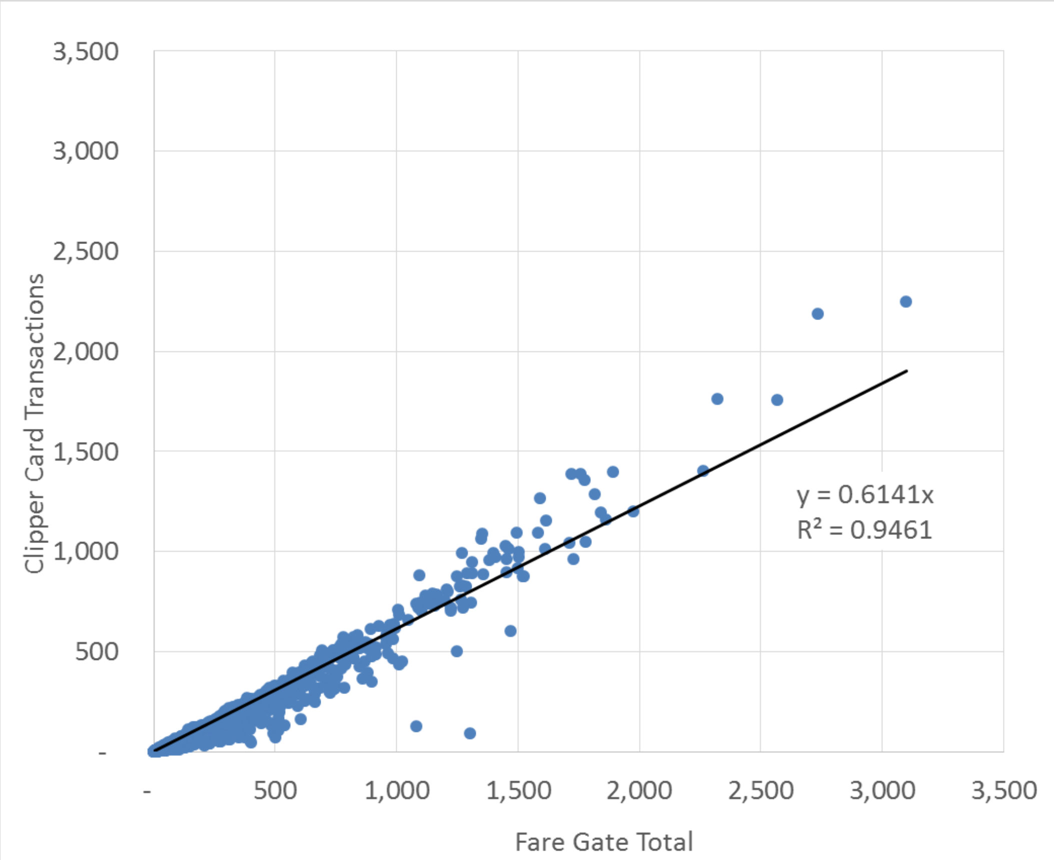


Figure 2.2: Trips between BART station pairs estimated from Clipper Card transactions and fare gate totals

which boardings are transfers and which are originating boardings.

As described in section 3.2, a Clipper tag-on that occurs within 90 minutes of a previous tag-on by the same card is considered a transfer. This value corresponds to valid free transfer window on MUNI, and reflects the difference in times between tag-ons. This is a very simplistic rule, but it serves as a starting point for the analysis. Varying time windows could be tested, and where transaction location data or tag-offs are available, those factors can be considered as well.

Table 2.4 shows the number of transfers per linked trip, as derived from the Clipper data. Table 2.5 shows the equivalent measures from the available onboard surveys. The Clipper transfer rates are generally lower than those recorded in the onboard surveys, which is interesting because at least one onboard survey found that trips involving a transfer are more likely to be paid for with a Clipper Card [145]. It is possible that the 90 minute time window is too short for some trips once in-

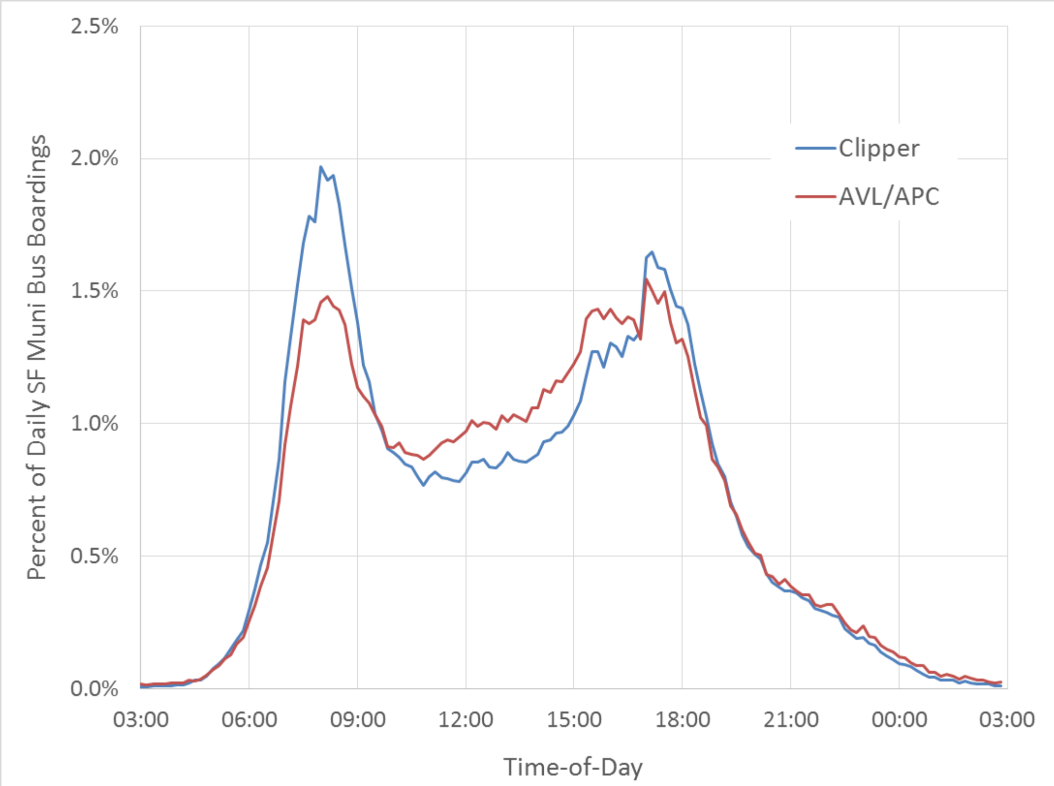


Figure 2.3: Time-of-day distribution from Clipper transactions and from AVL/APC data on MUNI buses

vehicle time and waiting time elapse between tag-ons. Alternatively, it may be that travellers are less likely to tag-on for a transfer than for an initial boarding.

Table 2.4: Transfers per linked trip, from Clipper transactions

		Linked Clipper Card Trips			
		Number of Transfers to Complete Trip			Boardings per Trip
Agency Name	Mode	0	1	2+	
AC Transit	Bus	68%	25%	7%	1.38
BART	Rapid Transit	89%	9%	2%	1.12
Caltrain	Commuter Rail	83%	15%	2%	1.19
Golden Gate Transit	Bus	88%	11%	2%	1.14
	Ferry	96%	4%	0%	1.04
MUNI	Bus	64%	27%	9%	1.45
	Cable Car	75%	19%	6%	1.31
	Light Rail	77%	19%	4%	1.27
SamTrans	Bus	68%	25%	7%	1.38
VTA	Bus	55%	32%	13%	1.58
	Light Rail	70%	23%	8%	1.38
SF Bay Ferry	Ferry	91%	9%	0%	1.10

Table 2.5: Transfers per linked trip, from onboard surveys

		Linked Clipper Card Trips			
		Number of Transfers to Complete Trip			Boardings per Trip
Agency Name	Mode	0	1	2+	
AC Transit	Bus	56%	36%	9%	1.55
Caltrain	Commuter Rail	67%	27%	6%	1.40
Golden Gate Transit	Bus	70%	22%	7%	1.35
	Ferry	84%	15%	1%	1.17
SamTrans	Bus	58%	33%	9%	1.51
SF Bay Ferry	Ferry	81%	17%	2%	1.21

2.4.6 Stability over Time

An important feature of smart card data is that they are collected continuously over time, providing the opportunity to monitor trends or measure the effects of transit system changes. To be useful for this purpose, though, they must measure the underlying system, and not an arbitrary data change.

To evaluate the usefulness of the data for this purpose, the average weekday Clipper transactions on BART and MUNI bus are plotted against fare gate or APC data for the same months, where available. These plots are shown in Figure 2.4 . The months of April 2013, August 2013 and February 2014 are not in the available Clipper database, but the surrounding months give a sense of the trend. The dip in ridership in July and October 2013 is due to strikes on the BART system. The Clipper data track the ridership moderately well, although with a bit more noise. The Clipper data for both systems show the dip for the strikes, for example. The rise in Clipper transactions in the final three months of data is interesting, because both MUNI bus and BART show a parallel rise, and it is faster than the growth in total BART ridership. One possibility is that the Clipper Card penetration increased at this time, which highlights the risk of assuming that an increase in card transactions is the same as an increase in ridership.

An appropriate response to this issue would to calculate a weight that varies with time. In fact, based on the analysis contained in Chapter 2, this is what is done in Chapter 3 for subsequent reporting of MUNI performance. As described at the end of Section 3.3.2, an additional component is added to the Clipper weight for

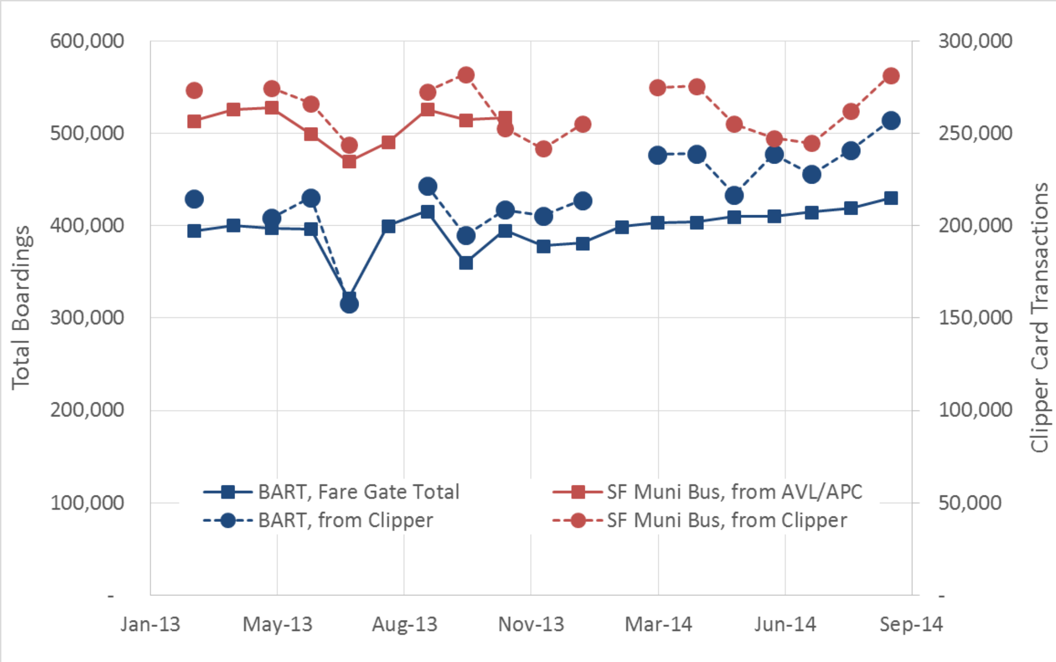


Figure 2.4: Trend in Clipper tag-ons on BART and MUNI, compared to total boardings

MUNI records that scales the weighted Clipper records to match the total boardings as determined by the weighted and expanded AVL-APC data. This applies in cases where both are available.

2.5 Estimated Model Correction Factors

In this section, the biases identified above are explored in a more systematic way, and a method is proposed to mitigate those biases. The basis for both is a discrete choice model of Clipper Card use estimated from the onboard survey data.

2.5.1 Estimated Models of Clipper Card Use

The models are binary logit models of whether an unlinked transit trip is made with a Clipper Card, or via another payment method. The model takes the form:

$$Pr(Clipper) = \frac{e^U}{1+e^U} \tag{2.1}$$

where $Pr(Clipper)$ is the probability that the unlinked trip is made using a Clipper Card, and U is the utility associated with using a Clipper Card. The utility can be expressed as $U = \beta X$ where β is the vector of estimated model coefficients and X is the vector of predictors. The model coefficients are estimated from the onboard survey data using maximum likelihood estimation. The use of onboard survey data is necessary because it provides a complete picture of trips made both with and without Clipper. The structure of the data is such that there is one record for each unlinked transit trip (boarding) in the surveys. If a Clipper Card was used to pay when the boarding is made, then the dependent variable is coded with a one, otherwise it is coded with a zero. Data are used for the six operators where the surveys are available. Additional onboard surveys are underway, which will allow for a more complete analysis in the future.

The estimated models allow for statistical inference of the factors correlated with Clipper use. In the results presented, t-statistics greater than 1.96 or less than -1.96 indicate the variable has a statistically significant correlation with Clipper use at the 95% level. If Clipper use was representative of the system as a whole, all of the descriptive variables would be expected to be insignificant. As the results in Table 2.6 show, this is not the case.

Table 2.6 shows the model estimation results for two models. Model A includes a full set of trip characteristics and socio-economic variables as available in the onboard surveys. It provides the best possible fit, and through the coefficient estimates shows the ways in which Clipper use is biased. Model B is limited to those variables that are available in the Clipper data. It is still estimated from the onboard surveys, but can be applied to the Clipper data to calculate correction factors, as will be discussed in Section 2.5.2. When a coefficient changes noticeably between Model A and Model B, it indicates that the variable is correlated with a variable that has been excluded. The hope is that the variables remaining in Model B are sufficiently correlated with the excluded socio-economic variables to allow Model B to partially or wholly adjust for the biases in those socio-economic dimensions.

The model results first show a constant associated with each operator. In addition, Clipper use is positively correlated with the use of express bus and BART, which can appear at any point during the linked trip. Trips with transfers are more likely to be made with a Clipper Card, likely because the cards provide a convenient

Table 2.6: Estimated models of Clipper Card use

Variable	Model A coefficient	Model A t-statistic	Model B coefficient	Model B t-statistic
Operator				
AC Transit	-0.758	-10.12	-0.396	-10.75
Caltrain	-0.763	-8.84	-0.186	-4.89
Golden Gate Transit (bus)	-1.608	-12.61	-1.164	-10.87
Golden Gate Transit (ferry)	0.198	1.26	1.004	7.46
SF Bay Ferry	-1.942	-12.75	-1.102	-8.64
SamTrans	-0.945	-11.58	-0.559	-11.63
Path Attributes (base=Other mode)				
Linked trip includes express bus	0.747	9.98	1.096	15.49
Linked trip includes BART	0.232	3.82	0.369	6.33
Transfers (base=None)				
Linked trip has 1+ transfers	0.133	3.66	0.161	4.67
Time-of-Day (base=Off-peak/missing)				
AM Peak	0.269	6.72	0.568	15.13
PM Peak	0.289	7.25	0.531	14.09
Fare Category (base=Adult)				
Youth	-0.186	-2.03	-0.113	-1.78
Senior	-0.198	-1.73	-0.164	-2.47
Disabled	-1.636	-18.39	-1.847	-21.69
Easypass or class pass	-0.815	-10.40	-0.622	-8.86
Other discount	-4.288	-9.36	-4.014	-8.78
Missing/do not know	-0.172	-0.45	-0.054	-0.15
Access Mode (base=Walk access)				
Linked trip has drive access or egress	-0.110	-2.33		
Tour Purpose (base=Other)				
Work	0.745	15.15		
College/university	0.599	8.69		
Grade/high school	0.872	8.06		
Household Income (base=under $10,000)				
$10,000 to $25,000	0.128	2.09		
$25,000 to $35,000	0.268	3.95		
$35,000 to $50,000	0.472	6.53		
$50,000 to $75,000	0.583	8.21		
$75,000 to $150,000	0.746	10.98		
$150,000 or higher	0.904	11.38		
Under $35,000 (if refused detailed income)	0.230	1.68		
$35,000 or higher (if refused detailed income)	0.695	5.19		
Missing/refused	0.098	1.35		
Race (base=White/Asian)				
Black	-0.187	-4.06		
Other	-0.237	-4.48		
Missing	0.023	0.17		
Hispanic (base=No)				
Yes	-0.215	-4.02		
Missing	-0.518	-2.39		
Age Group (base=25-64)				
Under 25	-0.195	-4.29		
65 or older	0.198	1.72		
Missing	0.380	2.95		
Auto Ownership (base=1+ autos)				
Zero autos in household	0.180	4.56		
Work Status (base=Non-worker)				
Full- or part-time	-0.146	-2.74		
Missing	-3.982	-7.42		
Student Status (base=no/missing)				
Student	0.194	3.51		
Model Statistics				
Observations	19080		19080	
Log-Likelihood at 0	-13224		-13224	
Log-Likelihood	-11617		-12237	
Rho-squared	0.122		0.075	

means of transferring between operators. There is no discount provided for inter-operator transfers, it merely avoids the necessity of carrying two separate tickets purchased from separate machines or websites. Trips made during the AM or PM peaks are also more likely to use Clipper. This is similar to what is observed in Figure 2.3, and may be related to tour purpose, as discussed below. The fare category provides a proxy measure of select socio-economic attributes, and is therefore important to calculating the correction factors. It is an imperfect proxy, because it is also related to the policies of the operator, such as the fare media upon which the discounts are applied and distributed.

The remaining variables are included only in Model A. In several cases a "Missing" category is retained in the model even though it is statistically insignificant. This is done in order to avoid biasing the remaining coefficients in that group. Drive-access trips are less likely to use Clipper, all else being equal. Work, university and school trips are all more likely to be made with Clipper than other trips. This is likely because those trips tend to happen on a daily basis, so travellers are more willing to make the effort to obtain a Clipper Card, than for trips that happen more infrequently. Consistent with Section 2.4.2, there is a clear income trend where high-income travellers are more likely to use Clipper. A similar outcome is found with ethnic and racial minorities, where blacks, other races and Hispanics are all less likely to use Clipper than whites and Asians. In contrast to prior expectations, persons aged 65+ are more likely to use Clipper, although that result is marginally insignificant. Persons under 25 less likely, when controlling for the other factors in the model. Similarly, having zero autos in the household is positively correlated with Clipper use. Workers are somewhat less likely to use Clipper and students are somewhat more likely.

The additional variables result in a better fit for Model A, which has a rho-squared of 0.122 compared to 0.075 for Model B. These rho-squared values are calculated with respect to a model with a single constant. For comparison, the log-likelihood of a model with a separate constant for each operator is -13059. If the rho-squared values were calculated with respect to this model instead, they would be 0.110 for Model A and 0.063 for Model B. Either way, the variables beyond the operator-specific constants describe a significant portion of the choice.

Collectively, these model results highlight some of the biases in the use of Clipper. Some of the correlations were expected, such as with trip purpose and income, but several would not otherwise be obvious. Of particular importance are the trends estimated for income and race, because it means that using the Clipper data unadjusted would under-estimate low-income and minority travel—a problem from environmental justice analysis.

2.5.2 Application of Correction Factors

The above model can be used to calculate the probability that an unlinked trip is made using a Clipper Card. If that probability is applied to all records in the onboard survey, it will provide an estimate of the trips made by Clipper. Conversely, if the reciprocal of that probability is applied to only the Clipper records, it will provide an estimate of the total trips. Therefore, that reciprocal probability is proposed as a correction factor to adjust for the biases in the use of Clipper. The correction factor is calculated as:

$$CF = \frac{1}{Pr(Clipper)} = \frac{1+e^U}{e^U} \tag{2.2}$$

where CF is the correction factor.

To demonstrate its application, the correction factors derived from Model A and Model B were applied to the Clipper records in the onboard surveys. In Table 2.7 and Table 2.8, these are compared to the total trips and the Clipper trips in the onboard surveys, with no correction factors applied. In all cases, the summaries do account for the survey weights.

Table 2.7 shows the boardings (unlinked trips) by operator. The surveys contain a total of 278,000 boardings, of which 140,000 use a Clipper Card. The correction factors are applied only to the Clipper records, and scale up the 140,000 trips they represent to an estimate of the total. Both models slightly over-estimated the total, with notable over-estimates for Golden Gate Transit buses and for SamTrans.

One possible explanation for this over-estimation is bias resulting from the asymmetry of the inverse transformation. Very small probabilities will result in correction factors that are very high, whereas very large probabilities will result in correction factors close to one. On average, this will tend to bias the correction factors upwards.

Table 2.7: Correction factors applied to onboard survey data, totals by operator

	Values				Percent Difference		
Operator	Total Trips	Clipper Trips	Model A	Model B	Clipper Trips	Model A	Model B
AC Transit	158,571	75,025	154,740	151,673	-53%	-2%	-4%
Caltrain	52,463	31,038	51,743	54,387	-41%	-1%	4%
Golden Gate Transit (bus)	11,986	6,933	14,562	13,731	-42%	21%	15%
Golden Gate Transit (ferry)	6,795	5,490	6,663	6,767	-19%	-2%	-0%
SF Bay Ferry	4,610	1,623	4,424	4,548	-65%	-4%	-1%
SamTrans	44,108	19,871	56,759	53,158	-55%	29%	21%
Total	278,533	139,981	288,890	284,263	-50%	4%	2%

Table 2.8 examines the share of trips made for the groupings of variables included in Model A but not in Model B. In the model estimation, these variables were found to be significantly correlated with Clipper use, which means that the unadjusted Clipper trips are a biased estimate of the total shares. The values shown in the Clipper Trips column are consistent with this expectation. The Model A correction factors are expected to do a reasonably good job of correcting for these biases, and the results show that the shares with the Model A correction are generally within a few percent of the total shares. The question is whether the Model B correction factors, which contain only variables that can be applied to the Clipper transaction data, produces shares more similar to Model A or to the Clipper Trips column.

The Model B results in Table 2.8 are somewhere in between the uncorrected values and Model A. The bias in drive-access shares is well corrected. The tour purpose shares still show a 16% under-estimate for the other purpose, but this is improved from a 32% under-estimate, and the share for the work purpose is also improved. Similarly, there remains an under-estimate of low-income trips and an over-estimate of high-income trips, but it is about half the magnitude found in the unadjusted results. The racial and Hispanic biases are improved, but not eliminated.

Table 2.8: Correction factors applied to onboard survey data, shares by category

Attribute	Values				Percent Difference		
	Total Trips	Clipper Trips	Model A	Model B	Clipper Trips	Model A	Model B
Drive Access Shares							
Yes	24%	28%	23%	24%	18%	-1%	2%
No	76%	72%	77%	76%	-6%	0%	-1%
Total	100%	100%	100%	100%	0%	0%	0%
Tour Purpose Shares							
Work	54%	63%	51%	57%	18%	-5%	6%
College/university	12%	11%	12%	13%	-5%	2%	7%
Grade/high school	5%	5%	5%	5%	6%	-4%	11%
Other	30%	20%	32%	25%	-32%	8%	-16%
Total	100%	100%	100%	100%	0%	0%	0%
Household Income Shares							
Under $10,000	14%	10%	14%	12%	-30%	4%	-15%
$10,000 to $25,000	17%	13%	18%	16%	-21%	6%	-6%
$25,000 to $35,000	11%	10%	11%	11%	-9%	3%	-2%
$35,000 to $50,000	9%	9%	9%	9%	7%	-1%	5%
$50,000 to $75,000	10%	12%	10%	11%	17%	-5%	7%
$75,000 to $150,000	17%	21%	16%	18%	27%	-5%	9%
$150,000 or higher	9%	12%	9%	11%	37%	-3%	16%
Under $35,000	2%	1%	2%	2%	-13%	9%	5%
$35,000 or higher	2%	2%	2%	2%	22%	-5%	14%
Missing/refused	10%	8%	10%	9%	-17%	-3%	-11%
Total	100%	100%	100%	100%	0%	0%	0%
Race Shares							
White/Asian	49%	56%	49%	53%	15%	0%	10%
Black	24%	21%	24%	22%	-13%	0%	-8%
Other	24%	21%	25%	22%	-14%	2%	-9%
Missing	3%	2%	2%	2%	-18%	-24%	-21%
Total	100%	100%	100%	100%	0%	0%	0%
Hispanic Shares							
Yes	21%	17%	21%	18%	-16%	3%	-11%
No	79%	82%	78%	81%	4%	-1%	3%
Missing	1%	1%	1%	1%	-15%	-3%	-16%
Total	100%	100%	100%	100%	0%	0%	0%
Age Group Shares							
Under 25	26%	24%	27%	25%	-9%	4%	-5%
25-64	65%	68%	64%	66%	5%	-1%	2%
65+	7%	6%	6%	7%	-13%	-4%	-0%
NA	2%	2%	2%	2%	11%	-6%	8%
Total	100%	100%	100%	100%	0%	0%	0%
Zero Auto Shares							
Yes	30%	27%	31%	30%	-11%	2%	-0%
No	70%	73%	69%	70%	5%	-1%	0%
Total	100%	100%	100%	100%	0%	0%	0%
Work Status Shares							
Full- or part-time	67%	75%	66%	70%	11%	-2%	3%
Non-worker	32%	25%	32%	30%	-22%	1%	-5%
Missing	1%	0%	2%	0%	-92%	149%	-93%
Total	100%	100%	100%	100%	0%	0%	0%
Student Shares							
Yes	29%	27%	30%	29%	-6%	4%	3%
No/missing	71%	73%	70%	71%	2%	-2%	-1%
Total	100%	100%	100%	100%	0%	0%	0%

The age shares are improved, particularly for the 65+ group, and the Model B corrections eliminate the bias against zero-auto households. It is interesting that the Clipper Trips column shows both persons aged 65+ and zero-auto households to be under-estimated in the unadjusted data, whereas the model coefficients on both those variables are positive. This demonstrates the complexity of the relationships in the data, and the importance of the model estimation. Both the work status shares and the student shares are corrected reasonably well.

With the exception of the two operators noted, Model A does a reasonable job of correcting of the biases of variables included in the model. Model B nearly eliminates the bias for some terms, and mitigates it for others. In spite of this improvement, important biases remain for tour purpose, income, and race/ethnicity.

2.6 Conclusions

In contrast to research conducted using relatively complete transit smart card data sets, the Clipper Card system in the San Francisco Bay Area is characterised by strong privacy restrictions and limited penetration rates. This research has evaluated the value of data with those limitations and proposed a method to develop correction factors for the data using choice models estimated from onboard survey data.

2.6.1 Data Evaluation

The data evaluation has focused on three key questions, as articulated at the start of Section 2.4. The conclusions to those questions are summarised here.

2.6.1.1 What New Information Do These Data Provide?

Perhaps the most significant advantage of smart card data, relative to other available data sources is that it provides data for a full day's worth of transit activity for each card. This allows transfers to be inferred, and trip linking and trip symmetry behaviour to be analysed. This is of particular importance in multi-operator, multi-modal transit region, where the local systems often serve as feeders to regional transit.

A second key advantage of the Clipper data is that it provides a common means of measurement across several different systems. For example, MUNI has an APC system on many of its buses to collect detailed passenger counts. The equipment is not installed on its light rail vehicles, so Clipper provides an alternative means of measuring ridership.

It would be quite valuable if the data could be used to derive transit origin-destination matrices throughout the system (as is possible on the BART), or at least the geographic distribution of boarding locations. This is where the data fall short for a combination of reasons. The Clipper Card technology does not record the location of a transaction. That location cannot instead be inferred from AVL data because the route ID and vehicle ID are not always coded, and because the exact time of the transactions are obfuscated for privacy reasons.

2.6.1.2 Are they Representative of the System as a Whole?

The analysis reveals that the Clipper data generally align with the geographic and temporal distribution of independent data sources where such sources are available, but that certain biases persist. Onboard surveys reveal a bias where high-income travellers are more likely to use Clipper. Geographic comparisons of BART data show that Clipper usage is higher for trips to the core area of San Francisco and Oakland than elsewhere in the region. Temporal comparisons show that the Clipper data overstate AM peak travel, and under-state midday travel.

These findings are important from an equity standpoint. If the planning process were to focus too heavily on meeting the needs of travellers as identified by smart card data, there is a risk of under-serving the needs of low-income and minority populations unless some mitigating steps are taken.

In addition, there is a risk in using smart card data over time that a changing penetration rate would be erroneously recorded as a change in ridership. To avoid this, it is beneficial to have external data, such as Automated Passenger Counter (APC) data for the purpose of calculating weights that vary over time. Alternatively, the risk is mitigated if the penetration rate is already high enough to be considered saturated.

2.6.1.3 What Can Be Done to Make the Data More Useful?

There may be some options to generate more complete bus boarding location information. One option would be to explore the possibility of obtaining technology that records the location of the transaction when the smart card system is next upgraded. Second, it might be possible to do the data matching on the agency side behind a firewall, and only release the processed data. Third, it might be possible to release more specific data to a limited number of users on a secure server, such as the Transportation Secure Data Center (TSDC) [150].

Within the existing framework, the most important piece of missing information is the route ID. On MUNI, for example, 94% of bus transactions have their route coded as "SFM Bus". The route ID can be provided using existing technology and without any loss of privacy, although existing labour agreements may be limiting.

2.6.2 Proposed Correction Factors

Onboard survey data were used to estimate binary logit models predicting the probability that a Clipper Card is used for each unlinked trip record. Two models were estimated: Model A with a full set of trip attributes and socio-economic variables and Model B with only those variables available in the Clipper transaction data. The estimation results confirm that the biases observed in the Clipper data are statistically significant in dimensions including access mode, tour purpose, income, race and ethnicity, age, auto ownership, worker status and student status.

The reciprocal of the modelled probability of using Clipper is proposed as a correction factor to be applied to Clipper transaction data. Applying these correction factors would make the Clipper data more representative of the system as a whole, although there will be some over-estimation due to the asymmetry in the use of the inverse function.

To demonstrate its application in comparison to known result, these correction factors are applied the Clipper transactions in the onboard surveys. This application shows that both models slightly over-predict the total ridership. Model A reasonably corrects the biases in the Clipper records for most operators, and Model B mitigates, but does not eliminate those biases. Important biases remain, particularly with respect to income and race/ethnicity. Therefore, users should be cautious if the data are used for equity or environmental justice analysis.

Chapter 3

Transit Data Fusion

This chapter describes the development of the core software tool used in this research. It fits within the larger system design, as described in Section 1.4 of the introductory chapter, bearing in mind the design goals enumerated there. This is a data intensive thesis, and a large portion of the effort is in developing the prototype software tools needed to combine those data in a way that is both accessible and representative of the system. That effort is described here, with the expanded data produced by the tool used in the subsequent chapters.

3.1 Introduction

Performance-based planning builds upon the traditional transport planning process by aligning planning goals and objectives with specific performance measures against which projects can be evaluated. The emergence of performance-based planning received a boost from the recent U.S. federal transportation legislation [32] which makes it more central to the overall planning process. In recent years, researchers and practitioners have made significant progress in developing approaches to performance-based planning [151], including approaches to establishing performance-based planning programs [152, 153], methods for converting data into performance measures [154, 155, 156], and experience formulating relevant performance measures from institutional priorities [152, 157, 153]. In spite of this momentum, a number of challenges still remain, including the availability of supporting data, the ability to synthesise those data into meaningful metrics and the resources required for analysis [158].

This research aims to meet these challenges by developing software tools to support the fusion and analysis of large, passively collected data sources for the purpose of measuring and monitoring transport system performance. Because they are continuously collected, Big Data sources provide a unique opportunity to measure the changes that occur in the transport system. This feature overcomes a major limitation of traditional travel data collection efforts, which are cross-sectional in nature, and allows for a more direct analysis of the changes that occur before-and-after a new transport project opens.

This work focuses on transit system performance, using San Francisco as a case study. It takes advantage of the Automated Vehicle Location (AVL) and Automated Passenger Counter (APC) data available on the city transit system, as well as data from the region's transit smart card system, Clipper Card.

As of the year 2000, automated data collection systems were becoming more common at transit agencies, but data systems were immature, network and geographic analysis methods were in their infancy, and the data were often used for little beyond federal reporting requirements [159]. Subsequently, Transit Cooperative Research Program (TCRP) Report 88 provided guidelines for developing transit performance measurement systems, with a focus on identifying appropriate performance measures to correspond to agency goals [160]. By 2006, TCRP 113 identified a wider range of AVL-APC applications, but still a dichotomy between APC data which was used in its archived form and AVL data which was often designed for real-time analysis and not archived or analysed retrospectively [81]. More complete data systems have since been developed that encapsulate the data processing and reporting [161, 84], apply data mining methods in an effort to improve operational performance [82], and examine bus bunching [79, 162]. Initial attempts have been made to visualise the data at a network level [163, 164].

Several characteristics distinguish this study from previous work.

First, it operates on a sample of AVL-APC data, and a methodology is established to expand the data to the schedule as a whole and weight the data to represent total ridership. This is in contrast to the examples given above which generally assume full data coverage. Establishing expansion and weighting methods is important because it allows Big Data analysis to be applied in a wider range of locations with lower expenditure on data collection equipment.

Second, the tool integrates transit smart card data. Others have demonstrated the value of smart card data either as a stand-alone data source [137, 138], or in combination with other data [139, 88]. Here, they are valuable because they are used to estimate the transfer rate, which cannot be determined from the AVL-APC data themselves.

Third, this study develops a tool to analyse the trends over a significant time period, from 2009 through the present, as opposed to many applications which focus on using the data to understand a snapshot of current operations [161, 162, 165, 166]. The tool allows data for any two time periods to be queried and compared at the analyst's request, and puts the focus specifically on the changes that occur in the system, and not just on observing current conditions. For example, changes that occur in a specific portion of the city may be traceable to housing developments or roadway projects at that location. These trends may go unnoticed given only aggregate measures or cross-sectional totals.

The remainder of this chapter is structured as follows: Section 3.2 describes the data sources used in this study. Section 3.3 covers the methodology for data processing, including the approach used to expand and weight the data to be representative of the system as a whole. Section 3.4 presents example outputs to demonstrate the types of performance reports that the data mashing tool can produce. Section 3.5 is conclusions and expected future work.

3.2 Data Sources

This research uses three primary data sources provided by the San Francisco Municipal Transportation Agency (SFMTA) and the Metropolitan Transportation Commission (MTC): AVL-APC data, archived General Transit Feed Specification (GTFS) data, and transaction data from the Clipper Card transit smart card system.

All three focus on the San Francisco Municipal Railway (MUNI) bus system in San Francisco.

The AVL-APC data is formatted with one record each time a transit vehicle makes a stop. At each stop, the following information is recorded:

- Vehicle location;
- Arrival time;
- Departure time;
- Time with door open;
- Time required to pullout after the door closes;
- Maximum speed since last stop;
- Distance from last stop;
- Passengers boarding;
- Passengers alighting;
- Rear door boardings;
- Wheelchair movements; and
- Bicycle rack usage.

In addition, identifiers are included to track the route, direction, trip, stop, sequence of stops, and vehicle number. The vehicle locations reflect some noise, both due to Global Positioning System (GPS) measurement error and due to variation in the exact location at which the vehicle stops. However, because the stop is identified, those locations can be mapped to the physical stop location, providing consistency across trips. The count data become less reliable as the vehicle becomes more crowded, but the data are biased in a systematic way, and SFMTA makes an adjustment in the data set to compensate for this bias. The data are not currently available on rail or cable car, only on the buses. Equipment is installed on about 25% of the bus fleet, and those buses are allocated randomly to routes and drivers each day at the depot. These data are available from 2008 to the present.

Because the AVL-APC data are available for only a sample of bus trips, the GTFS data are used to measure the scheduled universe of bus trips. GTFS is a data specification that allows transit agencies to publish their schedule information in a standard format. It was initially used to feed the Google Maps transit routing, and is now used by a wide range of applications. The data are in a hierarchical format and provide the scheduled time at which each vehicle is to make each stop. The full specification is available online [92]. The data used in this study were obtained from the GTFS archive [93], from 2009 to present.

The Clipper Card data provide information on fare transactions made with the cards. Clipper Card was introduced in 2010, and is used by eight transit operators in the Bay Area. The data are subject to California's laws governing personally identifiable information [167], making data privacy and protection issues of particular importance. Therefore, they have been released with a multi-step anonymisation and data obfuscation process [141]. Chapter 2 discusses the limitations of these data in detail. In spite of those limitations, the data provide value over the above other sources because they allow transfers to be identified.

For each transaction in the Clipper database, the following fields is available:

- Transaction Year;
- Transaction Month;
- Transaction Day of Week;
- Anonymised Clipper Card ID;
- Day Trip Sequence;
- Transit Agency;
- Payment Product;
- Fare;
- Tag-on Time;
- Tag-on Location;
- Route;

- Tag-off Time; and
- Tag-off Location.

As part of the data obfuscation process, the exact date is not provided, only the year month and day of week of the transaction. The Clipper Card ID is an anonymous ID that persists for one day. This means that it is possible to identify all transactions made by one card over the course of a single day, but not to identify transactions made by the same card over multiple days. The day trip sequence is the order of transactions by that card, over the course of the day. The transit agency is the agency on which the transaction is made, keeping in mind that a single linked trip can involve transfers between operators. The payment product identifies whether it is a cash fare or a pass, as well as the type of pass and any relevant discounts. The fare is the amount deducted for the transaction, which can be zero if there is a pass or transfer. The tag-on time is grouped into a 10 minute window, as is the tag-off time where it is relevant. The tag-on location is only known if it is a fixed location, such as a rail station, not if the tag-on occurs on a moving vehicle. The same applies to the tag-off location. The route should be recorded for all transactions, but most are actually missing for MUNI. This is because the driver must manually enter that information at the start of the shift, and not all do. Tag-offs are only relevant for systems with distance or zone based fares, such as Bay Area Rapid Transit (BART) and Caltrain, not to MUNI where the fare is only paid at boarding.

Clipper Card currently has a penetration rate of approximately 50% of riders on MUNI buses, meaning that only half of the actual boardings are recorded in the Clipper database. In addition, as part of the privacy protections, the data made available are a sample of 50% of the Clipper transactions, so together they represent about 25% of total ridership. While disaggregate transaction records are provided, the data are most useful at the aggregate level. This is because they cannot be located precisely in time, due to the obfuscation of the exact travel day, or in space, due primarily to the missing route information. What they can provide is average weekday estimates for the MUNI bus system as a whole, for each month.

3.3 Methodology

This section describes the methodology used to generate transit performance reports from the raw data. To ensure the performance measures are a valid representation of the transit system, the data area cleaned, expanded and weighted as outlined in Figure 3.1. In this figure, the Clipper data are referred to generically as transit smart card data.

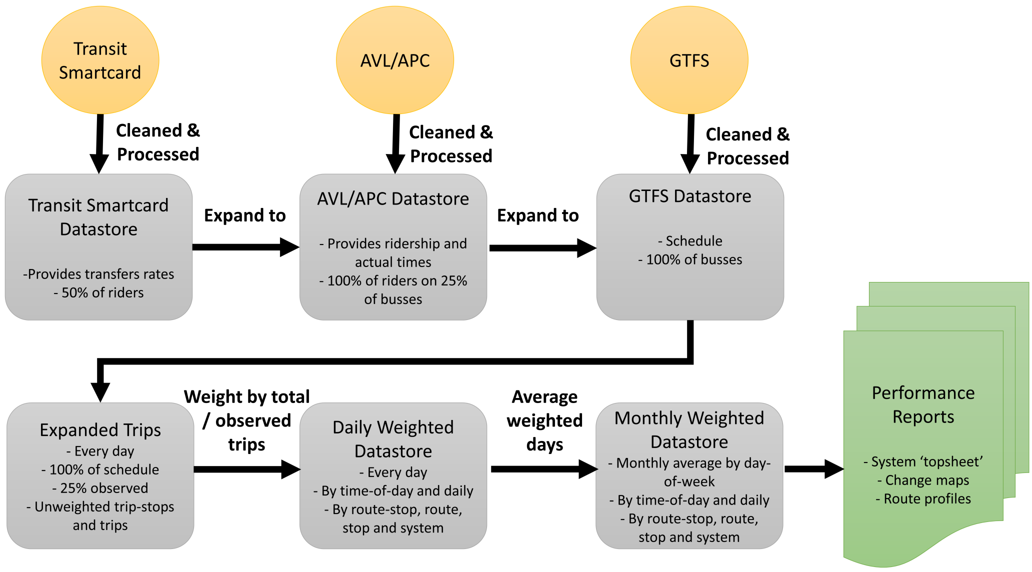


Figure 3.1: Data processing flow

3.3.1 Cleaning Individual Data Sets

First, each individual data set is cleaned and converted into a common format. For the AVL-APC data, this involves filtering out non-revenue service, records without a valid route ID, stop ID or trip ID, duplicate records, and those that do not meet quality control requirements. A number of derived fields are added, including the arriving and departing passenger load, the schedule deviation, flags for on-time arrival, time period groupings, and end-of-line flags. All date and time fields are converted from string format to a native Datetime format that allows for easy sorting and calculation of differences. An advantage of this class is that it supports normal arithmetic operations, such that subtracting a start time from an end time will give the correct difference, with the details of time zones, Daylight Savings

Time and leap years all fully encapsulated. As part of this Datetime conversion, special care is taken to handle the wrap-around effects of trips occurring between midnight and 3 am, which continue the schedule of the day prior, and whose ridership is counted with the day prior. An equivalency file is read to attach route IDs consistent with the GTFS data so the two files can later be joined. As part of this cleaning and processing, the data are converted from their raw text file format to an Hierarchical Data Format 5 (HDF5) Datastore format, as described later in this section. Throughout this chapter, when a Datastore is referenced, it is to this specific file format, as opposed to a more generic use of data set.

The raw GTFS data are read and converted to a record-based format such that they are directly comparable to the AVL-APC data. This format has one record for each stop made by each vehicle trip on each transit route. These data are written separately for each day, making the identification of weekday, Saturday or Sunday/holiday service explicit. The process makes time periods, trip IDs, direction IDs and route IDs consistent with the equivalency used for the AVL-APC data. It calculates the scheduled headway of each trip, the scheduled runtime from the previous stop, and the distance travelled from the last stop, and along the route shape as a whole.

The Clipper Card Data are processed to use consistent month and day of week codes as the other data. The mode is identified based on the agency name and the route name. This allows the MUNI bus transactions to be separated from the MUNI rail and cable car transactions, as well as from other modes and operators. A flag is added to identify transfers. If two subsequent tag-ons occur within 90 minutes of each other, the interchange is identified as a transfer. 90 minutes is selected based on the duration of time that MUNI allows passengers to board a second bus without paying additional fare. If there is a transfer, both the transaction before and after the transfer are identified as part of the same linked trip.

3.3.2 Data Expansion and Weighting

After the initial cleaning and conversion, the AVL-APC and GTFS data are joined to create an expanded Datastore. The goal of this expansion is to identify exactly what is missing from the sampled data, so they can be factored up to be representative of the universe as a whole. The relationship between the data sets is that transit smart card data provides a sample of about 50% of riders, the AVL-APC data provides 100% of riders on a sample of about 25% of vehicle trips, and the GTFS data identifies 100% of vehicle trips. Therefore, the expansion chain allows the more information-rich data sets to be combined with the more complete, but less rich data, much like a household travel survey would be expanded to match Census control totals. In this case, the expansion is a left join of the AVL-APC data records onto the GTFS records. Rail does not have AVL-APC equipment installed, so is excluded from the GTFS records as well. Note that this process is not able to account for scheduled trips that are not run, due to driver or equipment availability or other operational issues. The resulting Datastore has the full enumeration of service, but ridership and actual time information attached to only a portion of records. Without this step, it would not be possible to differentiate between trips that are missing because of a service change or those that are missing because they were simply not sampled. In a setting where we are explicitly interested in examining service changes, this distinction is important.

The output of this expansion process is a Datastore whose structure is shown in Table 3.1. The table also shows the source and data type of each field. These disaggregate records are referred to as trip-stop records because there is one record for each time a bus trip makes a stop (even if that stop is bypassed due to no passengers boarding or alighting). More specifically, records are defined by a unique combination of values in those fields identified with as an index in the source column. A related set of trip records is generated that aggregates across the SEQ field such that there is a single record for each time a bus makes a trip.

Table 3.1: Data dictionary for expanded and weighted trip-stop Datastore

Category	Field	Description	Type	Source
Time and Date	MONTH	Month and year	Datetime	Index
	DATE	Date	Datetime	Index
	DOW	Day of week (1=Weekday, 2=Saturday, 3=Sunday/Holiday)	Integer	Index
	TOD	Time of day	String	Index
Index Fields	AGENCY_ID	Agency ID (i.e. SFMTA)	String	Index
	ROUTE_SHORT_NAME	Route short name (i.e. 38)	String	Index
	ROUTE_LONG_NAME	Route long name (i.e. GEARY)	String	Index
	DIR	Direction (0=outbound, 1=inbound)	Integer	Index
	TRIP	Trip ID, as HHMM_SEQ of first stop on trip	Integer	Index
	SEQ	Stop sequence within route	Integer	Index
	ROUTE_TYPE	Type of route (0=tram, 3=bus, 5=cable car)	Integer	GTFS
Route Attributes	TRIP_HEADSIGN	Headsign on bus indicating destination (i.e. Ocean Beach)	String	GTFS
	HEADWAY_S	Scheduled Headway (min)	Float	Calculated
	FARE	Full fare ($)	Float	GTFS
	PATTERN	Pattern identifier calculated for all trips	String	GTFS
	PATTCODE	Pattern code (i.e. 38OB3)	String	AVL/APC
Stop Attributes	STOPNAME	Name of stop (i.e. Geary Blvd & Divisadero St)	String	GTFS
	STOPNAME_AVL	Name of stop in AVL/APC data	String	AVL/APC
	STOP_LAT	Latitude of stop location	Float	GTFS
	STOP_LON	Longitude of stop location	Float	GTFS
	SOL	Start of line flag (1=start of line, 0=not)	Integer	GTFS
	EOL	End of line flag (1=end of line, 0=not)	Integer	GTFS
	TIMEPOINT	Timepoint flag (1=stop is a timepoint in schedule, 0=not)	Integer	AVL/APC
Times	ARRIVAL_TIME_S	Scheduled arrival time	Datetime	GTFS
	ARRIVAL_TIME	Actual arrival time	Datetime	AVL/APC
	ARRIVAL_TIME_DEV	Deviation from arrival schedule (min)	Float	Calculated
	DEPARTURE_TIME_S	Scheduled departure time	Datetime	GTFS
	DEPARTURE_TIME	Actual departure time	Datetime	AVL/APC
	DEPARTURE_TIME_DEV	Deviation from departure schedule (min)	Float	Calculated
	DWELL_S	Scheduled dwell time (min)	Float	GTFS
	DWELL	Actual dwell time (min)	Float	AVL/APC
	RUNTIME_S	Scheduled running time (min), excludes dwell time	Float	GTFS
	RUNTIME	Actual running time (min), excludes dwell time	Float	AVL/APC
	TOTTIME_S	Scheduled total time (min), runtime + dwell time	Float	GTFS
	TOTTIME	Actual total time (min), runtime + dwell time	Float	AVL/APC
	SERVMILES_S	Scheduled service miles	Float	GTFS

Continuation of Table 3.1

Category	Field	Description	Type	Source
	SERVMILES	Service miles from AVL/APC data	Float	AVL/APC
	RUNSPEED_S	Scheduled running speed (mph), excludes dwell time	Float	Calculated
	RUNSPEED	Actual running speed (mph), excludes dwell time	Float	Calculated
	ONTIME5	Vehicle within -1 to +5 min of schedule (1=yes, 0=no)	Float	Calculated
Ridership	ON	Boardings	Float	AVL/APC
	OFF	Alightins	Float	AVL/APC
	LOAD_ARR	Passenger load upon arrival	Float	AVL/APC
	LOAD_DEP	Passenger load upon departure	Float	AVL/APC
	PASSMILES	Passenger miles	Float	Calculated
	PASSHOURS	Passenger hours, including both runtime and dwell time	Float	Calculated
	WAITHOURS	Passenger waiting hours, with wait as 1/2 headway	Float	Calculated
	PASSDELAY_DEP	Delay to passengers boarding at this stop	Float	Calculated
	PASSDELAY_ARR	Delay to passengers alighting at this stop	Float	Calculated
	RDBRDNGS	Rear door boardings	Float	AVL/APC
	CAPACITY	Vehicle capacity	Float	AVL/APC
	DOORCYCLES	Number of times door opens and closes at this stop	Float	AVL/APC
	WHEELCHAIR	Number of wheelchairs boarding at this stop	Float	AVL/APC
	BIKERACK	Bikerack used at this stop	Float	AVL/APC
Crowding	VC	Volume-capacity ratio	Float	Calculated
	CROWDED	Volume ¿ 0.85 capacity	Float	Calculated
	CROWDHOURS	Passenger hours when volume ¿ 0.85 capacity	Float	Calculated
Additional ID Fields	ROUTE_ID	Route ID in GTFS	Integer	GTFS
	ROUTE_AVL	Route ID in AVL/APC	Integer	AVL/APC
	TRIP_ID	Trip ID in GTFS	Integer	GTFS
	STOP_ID	Stop ID in GTFS	Integer	GTFS
	STOP_AVL	Stop ID in AVL/APC	Float	AVL/APC
	BLOCK_ID	Block ID in GTFS	Integer	GTFS
	SHAPE_ID	Shape ID in GTFS	Integer	GTFS
	SHAPE_DIST	Distance Along Shape (m)	Float	GTFS
	VEHNO	Vehicle Number	Float	AVL/APC
	SCHED_DATES	Dates when this schedule is in operation	String	GTFS
Weights	TRIP_WEIGHT	Weight applied when summarizing data at trip level	Float	Calculated
	TOD_WEIGHT	Weight applied when calculating time-of-day totals	Float	Calculated
	DAY_WEIGHT	Weight applied when calculating daily totals	Float	Calculated
	SYSTEM_WEIGHT	Weight applied when calculating system totals	Float	Calculated

The resulting Datastore has a record for every scheduled trip-stop, with a full enumeration of the associated schedule data, but ridership and schedule adherence information is only available on the 20-25% of records that are fully observed. For the remaining records, those fields are left with a missing value. The Datastore at this root level is suitable for making comparisons of individual trips or trip-stops. However, summing values across trips to generate time-of-day, daily, or system totals would result in an under-estimate of the total ridership because of the missing values. Therefore, a set of weights is developed to factor up the records to estimate the totals at these more aggregate levels for each day.

Because an entire trip is observed together, weights are calculated for trips and then broadcast to all stops in that trip. The weights are calculated by grouping the trips to the level of aggregation of interest, and within the group, applying the formula:

$$W_t = \frac{N}{\sum_t w_t} w_t \tag{3.1}$$

Where: W_t is the weight for trip t, N is the number of trips in the group, and w_t is the base weight for trip t. In cases where there are no observed trips (and therefore the denominator is zero), the resulting weight is set to zero, rather than an undefined value.

These weights are built hierarchically, such that the higher-level weights incorporate the lower-level weights. At the lowest level, a binary flag indicating whether or not the trip is observed is used as the base weight. The process is best explained by means of the example shown in Table 3.2.

The first set of columns shows basic attributes of the trip: ID, route, time-of-day and trip. The trip here is coded as the departure time from the first stop. For simplicity of the example, the combination of route and trip identify a unique record, although in reality it is the combination of date, agency ID, route, direction and trip. These data are available for every scheduled trip from the GTFS records.

The next column, Observed, is a binary flag indicating whether the trip is observed in the AVL/APC data. Only those records have ridership and actual (not

Table 3.2: Trip weighting example

ID	Route	TOD	Trip	Observed	Weights TOD	Weights Day	Weights System
1	1	AM	700	1	2	1.5	1.17
2	1	AM	800	0			
3	1	MD	1100	0	0		
4	1	MD	1400	0			
5	1	PM	1700	0	2		
6	1	PM	1800	1			
7	2	AM	730	0	2	1	
8	2	AM	830	1			
9	2	MD	1130	1	2		
10	2	MD	1430	0			
11	2	PM	1730	1	1		
12	2	PM	1830	1			
13	3	AM	800	0	0	0	
14	3	PM	1600	0	0		

scheduled) arrival departure times.

The last set of columns are the calculated weights, for three different levels of aggregation: time-of-day, day and system. The first two are specific to the route and the last is across all routes.

Within each TOD, the TOD weight is calculated as the ratio of scheduled trips to observed trips. The day weight is the ratio of scheduled trips to observed trips after the TOD weight has already been applied. For example, the day weight for route 2 is 1 because the TOD weighted observed trips already sum to 6, which is the number of scheduled trips for the day for route 2. The system weight is then calculated as the ratio of scheduled trips for all routes to observed trips after both the TOD and day weights have been applied. It can be verified that the product of the last four columns sums to 14, the total number of scheduled trips.

The different weighting levels allow the data to be tabulated appropriately for different levels of aggregation. For example, to obtain the total ridership on route 1 during the AM peak, the TOD weight would be applied to the observed records, but not the higher level weights.

The weights are assigned to the disaggregate records, and the data are aggregated with the weights applied to calculate route-stop, route, stop and system totals by time-of-day and for the daily total. This is done separately for each day, providing an estimate of the state of the system on each day for which data are available. The weighted data are then aggregated by month used to calculate conditions for an average weekday, an average Saturday and an average Sunday/holiday in each month. These monthly average Datastores are the primary source of information for the system performance reports, discussed in the next section, although the daily data remain available for more detailed analysis.

After calculating the average monthly conditions from the combined AVL-APC and GTFS Datastore, a separate set of weights is calculated for the Clipper records. There are three components to the Clipper weight. The first component simply scales up all records by a factor of two, to account for the fact that only a sample of 50% of the total Clipper records are included in the available data set. The second component converts from the monthly total to the average weekday, Saturday or Sunday conditions. It does this by dividing by the number of weekdays, Saturdays or Sundays in the data set for that month. The third component of the weight accounts for the fact that not all passengers pay their fare using a Clipper Card. Therefore the final weight is the ratio of total MUNI boardings from the expanded and weighted AVL-APC and GTFS Datastore to the total MUNI transactions in the Clipper data for that month. This weight is calculated separately for weekdays, Saturdays and Sundays, and then applied to all relevant Clipper records for that month. If more detailed Clipper data were available that fully identified the route or tag-on location, a more detailed set of geographically specific weights could be applied, but it the current format, it is limited to a system-level weight.

A second, linked-trip weight is then calculated by dividing the base weight by the number of individual transit legs on the linked trip. If there are no transfers then the divisor is one, if there is one transfer then the divisor is two, and so forth. From these data, linked trip totals and average transfer rates can be reported for each month and day of week type. These are reported in parallel to the aggregate monthly outputs of the AVL-APC and GTFS Datastore, even though the disaggregate records are only linked through the calculation of the weights.

3.3.3 Percent of Trips Observed

The estimates resulting from this process will be more reliable if there is reasonably good AVL-APC coverage of observations across routes. To examine the route coverage, Table 3.3 shows the percent of trips observed on each route for each weekday in July 2010. 22% of trips are observed, although this varies somewhat by route. The weighting process should do a good job of accounting for these varying penetration rates. More limiting are the cases where zero trips are observed on a route, which are highlighted with red cells. In these cases, the weighting process scales up the ridership on other routes to account for the missing values on that route. The missing values tend to occur on the routes that make fewer trips. Overall, 93% of routes are observed at least once during the month, with those routes covering 96% of trips.

One of the challenges in this effort is that the sampling of trips is not entirely random. There are operational constraints, such as certain types of buses (motor bus versus trolley bus, and articulated versus standard length) being needed on certain routes, and the fact that once a bus is assigned it tends to drive the same route back and forth. The result is that the data will not be as reliable as could be achieved with a well-designed sampling plan, but with a good overall coverage can be expected to provide good estimates of the state of the system.

To evaluate the magnitude of the error that can be expected from the sampling and weighting process, the number of service miles is used as an indicator. Service miles serves as a useful indicator because it is calculated from the GTFS data, so the enumerated value for the system as a whole is known. For comparison, the service miles are also calculated from the subset of observed records, with the weights applied to scale up those observed records to the system total. These calculations reveal that for months from 2009 through 2013, the average magnitude of the weighting error at the system level is 1.0%, and the maximum magnitude is 3.3%.

Table 3.3: Percent of trips observed on each route on weekdays in July 2010

Route		Scheduled Trips	Percent of Trips Observed on Date																					
			6/1	6/2	6/3	6/4	6/7	6/8	6/9	6/10	6/11	6/14	6/15	6/16	6/17	6/18	6/21	6/22	6/23	6/24	6/25	6/28	6/29	Average
1	CALIFORNIA	389	24%	20%	33%	13%	32%	15%	27%	35%	14%	30%	18%	23%	28%	19%	30%	30%	17%	25%	10%	20%	19%	23%
1AX	CALIFORNIA A EXPRESS	22	9%	32%	14%	23%	5%	9%	9%	18%	18%	27%	9%	18%	23%	32%	18%	0%	27%	27%	9%	18%	5%	17%
1BX	CALIRORNIA B EXPRESS	27	11%	11%	37%	22%	7%	7%	26%	11%	37%	11%	11%	7%	22%	7%	26%	26%	7%	22%	30%	26%	19%	18%
2	CLEMENT	118	15%	31%	26%	4%	37%	55%	23%	14%	29%	19%	27%	17%	31%	36%	28%	14%	15%	21%	21%	26%	25%	25%
3	JACKSON	128	18%	27%	11%	35%	0%	20%	59%	3%	47%	20%	0%	16%	45%	34%	6%	20%	16%	35%	48%	17%	38%	24%
5	FULTON	324	24%	26%	22%	31%	35%	20%	19%	25%	16%	14%	31%	24%	13%	19%	24%	34%	10%	26%	34%	20%	18%	23%
6	PARNASSUS	183	16%	17%	26%	21%	46%	16%	30%	42%	35%	18%	34%	52%	9%	28%	26%	33%	38%	23%	19%	26%	11%	27%
8AX	BAYSHORE A EXPRESS	42	24%	24%	26%	26%	17%	12%	31%	29%	12%	14%	26%	24%	29%	33%	5%	24%	33%	21%	29%	5%	26%	22%
8BX	BAYSHORE B EXPRESS	42	29%	33%	10%	12%	29%	36%	14%	14%	2%	21%	24%	29%	7%	36%	26%	24%	33%	19%	26%	12%	24%	22%
8X	BAYSHORE EXPRESS	175	25%	32%	5%	10%	33%	35%	12%	13%	0%	23%	26%	30%	5%	43%	21%	27%	29%	14%	29%	14%	24%	21%
9	SAN BRUNO	170	2%	13%	22%	26%	18%	26%	8%	20%	24%	9%	33%	25%	33%	11%	9%	24%	5%	14%	16%	22%	16%	18%
9L	SAN BRUNO LIMITED	119	5%	11%	6%	8%	0%	22%	11%	13%	10%	16%	10%	6%	0%	6%	23%	5%	10%	12%	10%	0%	34%	10%
10	TOWNSEND	87	7%	0%	0%	13%	6%	7%	0%	0%	20%	0%	20%	0%	10%	26%	7%	7%	6%	17%	7%	0%	14%	8%
12	FOLSOM/PACIFIC	102	13%	58%	36%	32%	15%	48%	19%	23%	0%	32%	7%	30%	13%	0%	8%	18%	13%	0%	35%	28%	0%	20%
14	MISSION	322	19%	20%	30%	19%	9%	25%	23%	25%	20%	15%	18%	24%	16%	14%	16%	21%	25%	24%	10%	22%	23%	20%
14L	MISSION LIMITED	124	34%	0%	29%	9%	8%	10%	23%	27%	19%	10%	8%	10%	15%	10%	40%	27%	19%	19%	31%	28%	18%	19%
14X	MISSION EXPRESS	33	12%	18%	30%	15%	24%	12%	33%	27%	39%	3%	15%	24%	33%	0%	9%	27%	18%	15%	24%	30%	6%	20%
16X	NORIEGA EXPRESS	29	31%	0%	41%	3%	38%	38%	7%	21%	7%	31%	38%	28%	14%	21%	21%	34%	3%	24%	10%	24%	34%	22%
17	PARK MERCED	64	0%	0%	0%	50%	0%	50%	50%	50%	0%	0%	50%	0%	50%	50%	0%	50%	0%	100%	0%	0%	0%	24%
18	46TH AVENUE	94	45%	29%	24%	20%	34%	29%	33%	56%	0%	6%	44%	47%	18%	26%	21%	41%	18%	83%	38%	39%	24%	32%
19	POLK	136	14%	27%	35%	18%	12%	47%	44%	24%	12%	51%	21%	35%	32%	31%	27%	17%	28%	33%	37%	30%	23%	28%
21	HAYES	174	5%	40%	32%	24%	13%	38%	29%	0%	52%	36%	27%	23%	52%	37%	28%	38%	19%	14%	2%	39%	26%	27%
22	FILLMORE	270	29%	23%	41%	26%	45%	32%	34%	19%	38%	31%	27%	24%	37%	42%	30%	22%	33%	22%	31%	31%	39%	31%
23	MONTEREY	108	21%	21%	19%	48%	13%	26%	20%	29%	6%	34%	13%	19%	16%	21%	13%	0%	17%	0%	31%	6%	13%	18%
24	DIVISADERO	204	15%	18%	30%	23%	19%	35%	23%	37%	30%	42%	46%	19%	35%	30%	24%	36%	26%	27%	46%	30%	50%	31%
27	BRYANT	139	0%	43%	9%	15%	17%	9%	29%	22%	19%	34%	6%	0%	5%	21%	0%	10%	19%	6%	14%	21%	30%	16%
28	19TH AVENUE	175	47%	24%	20%	13%	27%	18%	9%	44%	27%	13%	22%	57%	25%	38%	15%	32%	33%	26%	19%	30%	22%	27%
28L	19TH AVENUE LIMITED	49	51%	35%	10%	14%	33%	18%	8%	59%	29%	10%	27%	63%	12%	39%	18%	35%	24%	16%	16%	27%	29%	27%
29	SUNSET	177	11%	19%	29%	16%	12%	23%	33%	21%	30%	21%	12%	18%	12%	11%	29%	31%	18%	16%	33%	20%	17%	21%
30	STOCKTON	423	33%	22%	27%	39%	36%	19%	32%	30%	33%	42%	26%	39%	35%	46%	34%	42%	23%	46%	27%	37%	22%	33%
30X	MARINA EXPRESS	48	23%	13%	33%	13%	21%	29%	27%	17%	33%	40%	21%	25%	40%	21%	29%	21%	31%	19%	8%	27%	33%	25%
31	BALBOA	164	13%	26%	0%	13%	10%	43%	26%	23%	18%	20%	12%	37%	6%	12%	66%	0%	48%	39%	29%	34%	41%	25%
31AX	BALBOA A EXPRESS	21	5%	5%	29%	10%	10%	10%	5%	19%	24%	5%	10%	24%	19%	14%	19%	19%	10%	10%	5%	29%	5%	13%
31BX	BALBOA B EXPRESS	22	14%	23%	9%	14%	18%	23%	27%	18%	18%	14%	18%	18%	18%	23%	14%	18%	14%	14%	27%	18%	9%	18%
33	STANYAN	118	35%	29%	31%	17%	17%	15%	30%	18%	43%	17%	30%	52%	52%	35%	51%	17%	31%	66%	0%	63%	0%	31%

Continuation of Table 3.3

Route		Scheduled Trips	Percent of Trips Observed on Date 6/1	6/2	6/3	6/4	6/7	6/8	6/9	6/10	6/11	6/14	6/15	6/16	6/17	6/18	6/21	6/22	6/23	6/24	6/25	6/28	6/29	Average
35	EUREKA	60	0%	53%	95%	98%	62%	0%	0%	0%	63%	0%	0%	0%	20%	0%	0%	0%	75%	0%	0%	0%	98%	27%
36	TERESITA	64	33%	19%	0%	0%	19%	33%	0%	64%	22%	33%	8%	55%	33%	0%	34%	53%	58%	0%	0%	0%	31%	24%
37	CORBETT	91	0%	5%	19%	0%	2%	27%	0%	0%	0%	55%	51%	0%	0%	26%	0%	27%	27%	30%	22%	22%	43%	17%
38	GEARY	312	8%	8%	2%	31%	23%	11%	17%	16%	27%	34%	23%	12%	16%	9%	40%	13%	13%	38%	25%	14%	24%	19%
38AX	GEARY A EXPRESS	24	13%	4%	8%	13%	17%	4%	8%	13%	17%	13%	8%	25%	4%	4%	4%	4%	17%	21%	17%	0%	8%	11%
38BX	GEARY B EXPRESS	23	13%	26%	13%	13%	9%	22%	22%	17%	13%	13%	17%	13%	17%	9%	9%	13%	17%	22%	22%	13%	17%	16%
38L	GEARY LIMITED	276	17%	35%	12%	13%	18%	10%	13%	24%	12%	20%	12%	25%	14%	30%	12%	14%	21%	12%	8%	16%	15%	17%
39	COIT	62	0%	0%	52%	0%	48%	52%	0%	0%	98%	50%	48%	52%	0%	0%	52%	0%	48%	0%	48%	18%	39%	29%
41	UNION	96	40%	30%	45%	29%	40%	35%	15%	29%	36%	23%	43%	39%	41%	45%	32%	39%	28%	32%	33%	40%	33%	35%
43	MASONIC	160	31%	46%	45%	41%	16%	16%	36%	18%	34%	18%	19%	23%	38%	32%	42%	28%	19%	33%	51%	19%	29%	30%
44	O'SHAUGHNESSY	167	38%	29%	32%	17%	32%	19%	16%	19%	24%	13%	32%	14%	19%	45%	3%	26%	26%	25%	21%	6%	11%	22%
45	UNION-STOCKTON	200	17%	45%	1%	4%	22%	16%	4%	35%	4%	15%	9%	33%	14%	31%	25%	6%	35%	0%	25%	11%	5%	17%
47	VAN NESS	203	15%	5%	37%	27%	29%	8%	12%	11%	7%	64%	33%	7%	21%	31%	40%	45%	40%	36%	10%	26%	26%	25%
48	QUINTARA - 24TH STREET	147	33%	9%	28%	14%	24%	2%	18%	24%	20%	39%	14%	24%	18%	12%	25%	31%	22%	12%	7%	41%	5%	20%
49	MISSION-VAN NESS	289	28%	17%	17%	20%	22%	25%	21%	20%	20%	25%	12%	18%	19%	15%	20%	23%	9%	20%	26%	17%	10%	19%
52	EXCELSIOR	68	9%	0%	44%	44%	10%	9%	0%	31%	37%	44%	0%	9%	0%	0%	0%	43%	44%	0%	44%	0%	0%	18%
54	FELTON	104	38%	11%	16%	0%	40%	17%	5%	13%	23%	0%	16%	12%	16%	54%	3%	15%	0%	23%	0%	24%	31%	17%
56	RUTLAND	58	0%	69%	0%	0%	0%	0%	64%	0%	0%	0%	0%	0%	0%	59%	45%	0%	0%	21%	100%	66%	0%	20%
66	QUINTARA	93	0%	0%	47%	44%	0%	0%	45%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	53%	0%	9%
67	BERNAL HEIGHTS	96	0%	21%	17%	0%	0%	0%	0%	0%	0%	0%	27%	0%	79%	50%	0%	100%	0%	0%	16%	100%	50%	22%
71	HAIGHT-NORIEGA	141	17%	34%	16%	26%	14%	35%	26%	28%	16%	28%	27%	23%	12%	28%	23%	10%	25%	35%	6%	30%	24%	23%
71L	HAIGHT-NORIEGA LIMITED	25	12%	44%	20%	20%	20%	36%	20%	24%	16%	36%	44%	20%	12%	28%	28%	12%	32%	36%	8%	36%	28%	25%
80X	GATEWAY EXPRESS	1	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
81X	CALTRAIN EXPRESS	6	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
82X	LEVI PLAZA EXPRESS	22	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
88	B.A.R.T. SHUTTLE	16	0%	50%	50%	50%	50%	0%	50%	50%	50%	50%	0%	0%	50%	50%	19%	0%	50%	100%	0%	0%	0%	32%
90	OWL	13	0%	0%	0%	0%	0%	54%	0%	0%	46%	0%	54%	46%	100%	0%	0%	0%	46%	0%	0%	46%	0%	19%
91	OWL	15	40%	0%	0%	47%	20%	0%	0%	0%	27%	33%	40%	0%	20%	0%	13%	13%	13%	0%	47%	40%	67%	20%
95	INGLESIDE APTOS	2	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
108	TREASURE ISLAND	150	21%	28%	25%	33%	55%	16%	31%	63%	23%	31%	21%	28%	24%	29%	35%	0%	20%	7%	0%	7%	7%	24%
KM BUS		262	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
K-OWL		3	33%	33%	0%	0%	0%	0%	33%	0%	0%	33%	0%	0%	33%	33%	0%	0%	33%	0%	0%	0%	0%	11%
L-OWL		9	56%	56%	0%	0%	0%	0%	44%	0%	0%	44%	0%	0%	44%	44%	0%	0%	44%	0%	0%	44%	0%	18%
N-OWL		12	25%	0%	25%	0%	0%	0%	0%	25%	33%	33%	0%	58%	58%	0%	33%	33%	0%	17%	0%	0%	33%	18%
Total		8,092	19%	22%	23%	21%	22%	21%	21%	23%	21%	24%	21%	23%	22%	25%	23%	23%	21%	23%	21%	23%	21%	22%

3.3.4 Software Environment

The software was developed in an open-source framework in the Python environment. It is available under the GNU's Not Unix! (GNU) General Public License Version 3 for distribution [115]. In its current state, the software is a prototype that works specifically with San Francisco data, but it could be adapted for use elsewhere with moderate effort. It leverages several open-source packages specifically designed to provide high-performance data storage, access and analysis for extremely large data sets. Specifically:

1. Pandas is used for in-memory data operations, providing data structures and analysis tools for fast joins, aggregations, and tabulations of the data. Its functionality is similar to what is available in an R dataframe.

2. Hierarchical Data Format 5 (HDF5) is used to store the data on disk. It is designed for the fast and flexible storage of large data sets, allows for any combination of key-value pairs to be written, and allows on-disk indexing of the data.

3. PyTables is a package for managing hierarchical datasets designed to easily cope with extremely large data sets. PyTables serves as the interface between Pandas operations in memory and the HDF5 storage on disk.

The advantage to using this combination of technology is that it allows datasets too large to be stored in memory to be written to disk, but allows for random access to those data with very fast queries. The development has shown that the converted data are dramatically faster to access than in their raw text format. This workflow also provides much greater flexibility than using a traditional database, which typically perform best with a stable data structure, making them less ideal for exploratory analysis.

The process was tested in the Windows operating system on a machine with a dual-core 2.1 gigahertz processor and 16 gigabytes of memory. It runs as a single-stream process, and on this machine it took about two days to process the 2009 through 2013 data. The hard disk footprint of the resulting expanded and weighted

Datastores is about 400 gigabytes. This scale fits the definition of Big Data as "any data that cannot fit into an Excel spreadsheet" [168], although it is still practical to work with these data using standard hardware. It is of the scale, though, that software design decisions matter in terms of making the end result usable, and the ability to pre-calculate then query the most computationally intensive portions of the problem help this. Also, the software is structured that new data is appended to the end of the existing Datastore, so there is not a need to re-compute the whole as new data are added.

3.4 Sample Results

This section presents sample results from the data fusion tool. The purpose of this section is to illustrate the types of performance measure the tool is capable of reporting, and how those measures might be useful in planning. In all cases, the performance reports seek to report information that is both relevant to the planning process and readily explainable to policy makers. It further seeks to put the focus of the analysis on the changes that occur over time, rather than a single snapshot of the system.

Table 3.4 shows a sample of the monthly transit performance report. It consolidates the core performance measures onto a single page, and compares them to performance from another period, often the month before. The measures are grouped in the following categories:

1. Input Specification: Attributes selected by the user to define the scope of the report. The geographic extent can be the bus system as a whole, a route or an individual stop, with some minor differences for the route or stop reports. The day-of-week is weekday, Saturday or Sunday/holiday. Time-of-day can be specified for the daily total, or for individual time periods allowing for evaluation of peak conditions. The report generation date and a comments section are provided. The notes in this case indicate that system-wide service cuts occurred between the two periods, which corresponds to the nearly 10% reduction in service miles. The report is generated for the same month before

Table 3.4: Sample transit performance summary report

SFMTA Transit Performance Report

Input Specification

Geographic Extent:	All Busses
Day-of-Week:	Average Weekday
Time-of-Day:	Daily
Report Generated on:	2015-07-29 13:48:00
Comments:	Service cuts in April 2010

	Periods			
	Jul-2009	Jul-2010	Difference	% Difference
Service Provided				
Vehicle Trips	9,183	8,092	-1,091	-11.9%
Service Miles	57,751	52,046	-5,705	-9.9%
Ridership				
Boardings	526,423	509,400	-17,023	-3.2%
Rear-Door Boardings	1,688	1,436	-253	-15.0%
Passenger Miles	1,034,838	1,043,599	8,760	0.8%
Passenger Hours	123,675	125,622	1,947	1.6%
Wheelchairs Served	1,066	1,281	215	20.2%
Bicycles Served	1,791	1,713	-78	-4.4%
Level-of-Service				
Average Run Speed (mph)	10.60	10.29	-0.31	-2.9%
Average Dwell Time per Stop (min)	0.20	0.23	0.03	14.4%
Average Scheduled Headway (min)	13.84	14.34	0.51	3.7%
Average Full Fare ($)	2.00	2.00	0.00	0.0%
Average Distance Traveled per Passenger (mi)	1.97	2.05	0.08	4.2%
Average Passenger Speed (mph)	8.37	8.31	-0.06	-0.7%
Average Wait Time per Passenger (min)	5.34	5.70	0.37	6.9%
Reliability				
Percent of Vehicles Arriving On-Time (-1 to +5 min)	66.8%	63.0%	-0.04	-5.8%
Average Waiting Delay per Passenger (min)	2.74	2.49	-0.24	-8.9%
Average Arrival Delay per Passenger (min)	2.22	2.00	-0.22	-10.0%
Crowding				
Average Volume-Capacity Ratio	0.43	0.47	0.04	10.5%
Percent of Trips with V/C > 0.85	7.1%	8.8%	1.6%	22.8%
Passenger Hours with V/C > 0.85	7,143	8,411	1,268	17.8%
Observations & Error				
Number of Days	22	21	-1	-4.5%
Days with Observations	22	21	-1	-4.5%
Percent of Trips Observed	19.5%	22.1%	2.6%	13.2%

and after these cuts to report to avoid reporting seasonal changes.

2. Service Provided: The service provided metrics measure the total scheduled transit service, as found in the GTFS. Identical values mean that the schedule did not change between those two months.

3. Ridership: Ridership measures provide the total passenger boardings, the distance and time passengers spend onboard, and the number of wheelchairs and bicycles served. In this example, the ridership decreases by 3.2%, potentially in response to the service cuts.

4. Level-of-Service: The level-of-service section provides measures of the quality of service provided, as experienced by users. The average run speed, the dwell time per stop and the scheduled headway are measured as a function of the buses themselves. Run speed is defined as the speed between stops, so excludes the dwell time at stops. Scheduled headway is measured at each route-stop, calculated as the time from the previous trip of the same route. That is, it accounts for combined headways for multiple patterns of the same route, but it does not account for combined headways across multiple routes. The fare is reported as the average full cash fare across all routes and stops, as shown in the GTFS. Separate revenue data would be needed to measure the average fare paid accounting for discounts and passes. The average distance travelled, average passenger speed and average passenger wait are measured as a function of the passengers themselves. In contrast to the run speed, the average passenger speed includes dwell time, making it generally slower. Average waiting time is measured as half the scheduled headway, assuming random passenger arrivals and perfect reliability. The system-wide average passenger wait tends to be less than half the system-wide average scheduled headway because passengers tend to use more frequent service. In this example, both the average scheduled headway and passenger wait increase, which is logical given less frequent bus service.

5. Reliability: Reliability measures indicate how well the buses adhere to their schedule. Consistent with the Transit Capacity and Quality of Service Manual

(TCQSM) [169], a vehicle is considered on-time if it departs from a timepoint no more than one minute early or arrives more than five minutes late. In addition, two measures of delay are reported which are weighted to passengers instead of buses. The waiting delay is the average time passengers wait at their stop for a bus to arrive after its scheduled arrival time. Arrival delay is the average time passengers arrive at their alighting stop, past the scheduled time.

6. Crowding: For the purpose of this tool, a vehicle is considered to be crowded if the volume of passengers onboard exceeds 85% of the capacity. The range of 85-100% of total capacity corresponds roughly to the range of 125-150% of the seated load, which is referenced in the TCQSM as the maximum design load for peak-of-the-peak conditions. The crowding statistics report the average volume-capacity ratio, the percent of trips where the vehicle is crowded at some point during the trip, and the number of passenger hours in crowded conditions. These performance reports can easily be generated for each time period, allowing for monitoring of crowding during the peak periods.

7. Observations: The report includes the percent of trips observed, the total number of days and the number of days with observations. At a system level, there will generally be observations on each day, but specific routes or stops may not be observed on some days. The measurement error calculates the percent difference between the total boardings and alightings, providing an indication of the level of error that can be expected from the APC technology. The weighting error calculates the percent difference between the scheduled service miles and the weighted and expanded service miles, giving an indication of the error that can be expected as a result of the sampling and weighting process.

This performance report provides an overview allowing planners to quickly scan a range of indicators for changes that might be occurring.

While the numeric performance measures provide valuable information, their aggregate nature can wash out change that may be occurring in one portion of the city. Therefore, an interactive mapping tool was developed to plot key metrics in their geographic context. Figure 3.2 shows a screenshot from this tool. The left map shows a before period, the middle map an after period, and the right map shows either the absolute or relative change between the two periods. In this case, the comparison is between July 2009 and July 2010, before and after the service cuts in spring 2010. The user can select which time-of-day, which performance measure and which direction to plot. In this instance, the user has chosen to map the degree of crowdedness in the outbound direction during the 4-7 pm time period. The warm colours on the left two maps indicate more crowding, as measured by the average volume-capacity ratio during the period. The results are logical, with reasonably full buses moving west from the central business district towards residential areas of the city, as well as North-South on Van Ness Avenue. The map on the right shows the relative change in the metric between the two periods, with the warm colours indicating an increase in crowdedness and the cool colours indicating a decrease. In this instance, the change is concentrated on about three specific routes.

To accommodate further analysis of the changes that occur to specific routes, the software generates route profiles as shown in Figure 3.3. In this example, average weekday ridership on the 1-California route is plotted in the inbound direction during the AM peak. The x-axis is the sequence of stops along the route. The line charts show the number of passengers on the bus between each stop. The bar charts show the number of passengers boarding and alighting at each stop, with positive bars indicating boardings and negative bars indicating alightings. In all cases, the blue colours indicate the July 2009 period, and the red colours indicate the July 2010 period. The pattern of ridership remains similar between the two periods, with riders accumulating through the residential portions of the route, and passengers getting off the bus when it reaches the central business district, starting at the Clay Street and Stockton Street stop. The PM peak ridership profile would show the reverse. The route was shorted by the July 2010 period, with service no longer provided to the last three stops. Therefore, in the July 2010 period there are no alightings at these stops, and an increase in alightings at the new end-of-line stop.

The overall volume on this route during the AM peak is lower after these changes. These boarding profiles are useful when evaluating service changes made to specific routes, or the ridership resulting from newly opened land developments.

Finally, line plots are output, as in Figure 3.4, to show the trends over a longer period of time, rather than just for two periods. This particular example shows the on-time performance, defined as the share of buses arriving no more than one minute early or five minutes late. This is plotted for the daily totals, the AM peak and the PM peak. The results show the on-time performance is generally 60-70%, with higher values in the AM peak and lower values in the PM peak. Any of the performance measures can be easily plotted in this way, and doing so is an important step to understanding whether the changes observed are real, or simply within the natural variation of the data.

The software can automatically generate each of the performance reports described above, allowing for core analysis of the most important measures. In addition, the full weighted and imputed Datastore is available for advanced users who seek to conduct further in-depth analysis or custom queries. Files suitable for model estimation can also be written. The contents of those files are described in more detail in Chapter 4.

Figure 3.2: Sample transit performance change map

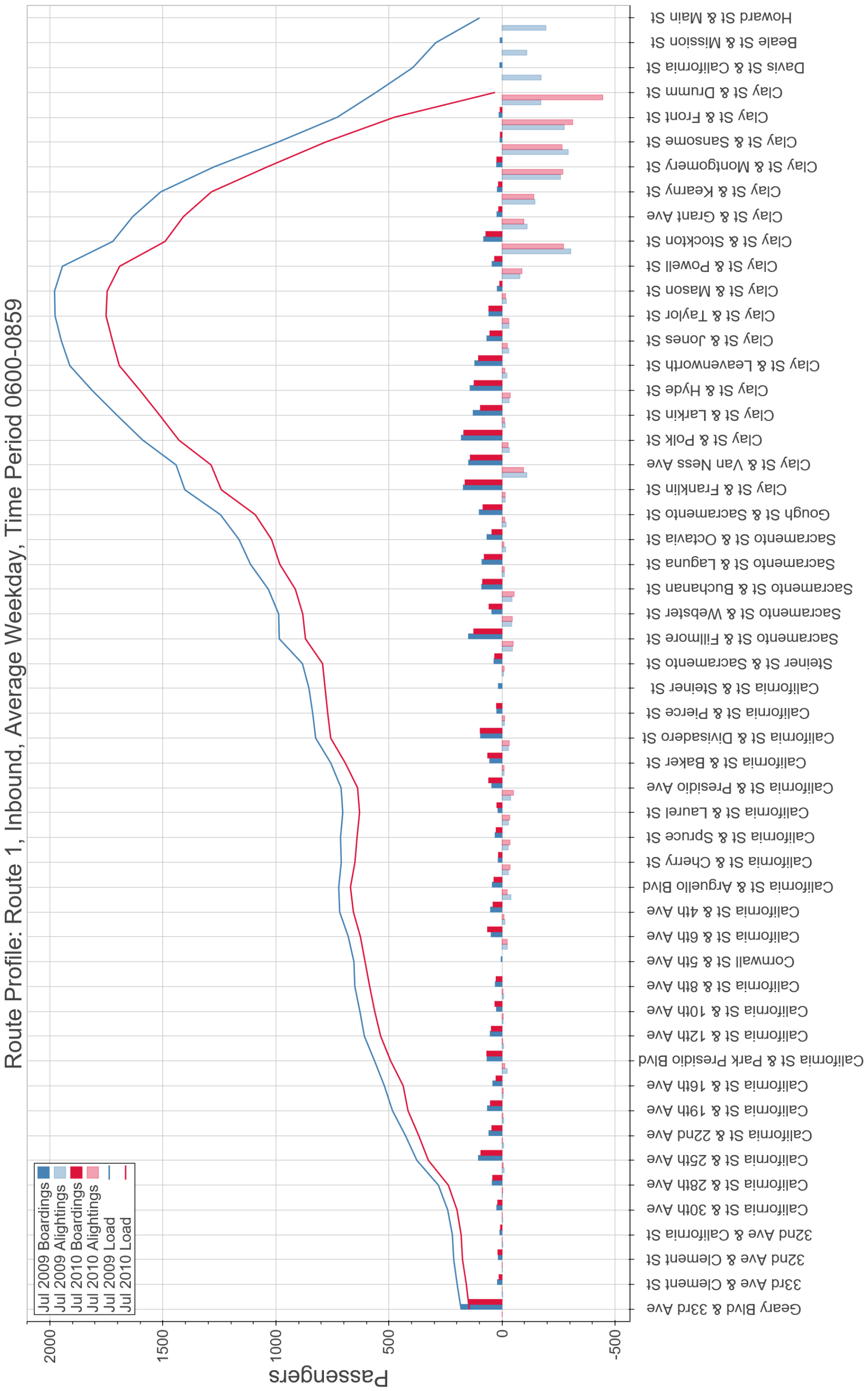


Figure 3.3: Sample route profile

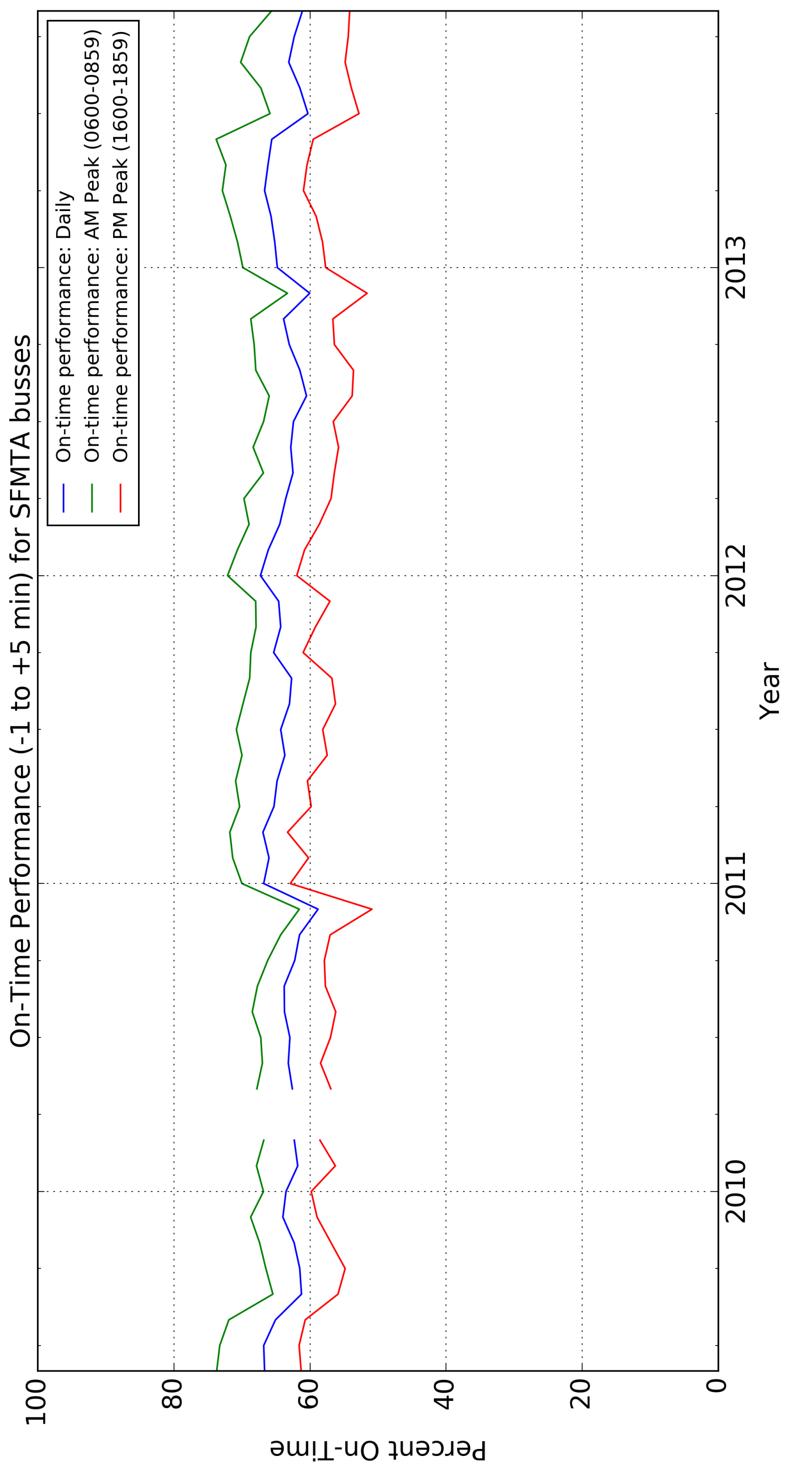


Figure 3.4: Sample trend plot

3.5 Conclusions and Future Development

The product of this research is a Big Data fusion tool that can be used to measure transit system performance over time. The software is implemented for San Francisco, but can be adapted for use in other regions with similar data.

The chapter addressed some of the methodological and mechanical challenges faced in managing these large data sets and translating them into meaningful planning information. One such challenge was the sampled nature of the data, where not all vehicles have AVL-APC equipment installed. To make these data more representative of the system as a whole, the vehicle trips in the AVL-APC data are expanded to match the universe of vehicle trips identified by the GTFS data and weights are developed to scale up to compensate for data that remain unobserved. The expansion process applies strategies from traditional surveys where a small but rich data set is expanded to match a less rich but more complete data set. Such strategies are key to spreading the use of Big Data for urban analysis beyond the first tier of cities that have near-complete data sets to those that are constrained by partial or incomplete data.

The software is available under an open-source licence [115]. For working with these large data sets, it was an important decision to work with libraries that allow fast querying of on-disk data, but also the ability to easily modify the data structure.

The data mashing tool reports and tracks transit system performance in the core dimensions of: service provided, ridership, level-of-service, reliability and crowding. The performance measures are reported for the system, by route and by stop, can also be mapped using an interactive tool. The focus of the tool is on providing the ability to monitor the trends and changes over time, as opposed to simply analysing current operations. By making performance reports readily available at varying levels of resolution, and the data mashing tool encourages planners to engage in data-driven analysis on an ongoing basis.

One future enhancement could be to incorporate automatic alerts for noteworthy changes. The software could monitor the data as they are accumulated and identify cases where the values change by more than a given percentage, or go beyond a set of predefined thresholds. When such a case is encountered, the software would send an alert to the user notifying her, and suggesting that they look at the situation more closely. This would further reduce the user burden to monitor a large number of indicators, and allow them to spend time on more interesting analysis.

The outputs of this tool are used in Chapter 4 to estimate time series models of transit ridership. Those models are subsequently applied to understand the drivers of transit ridership changes and to measure the ridership impacts of transit system changes.

Chapter 4

Time Series Model Estimation

Chapter 3 described the development of a Big Data mashing tool for measuring transit system performance. That tool combined several related and individually incomplete data sources, and weighted them appropriately to be representative of the San Francisco Municipal Railway (MUNI) bus system as a whole. The software included tools to visualise changes in transit performance, and create performance reports to summarise key indicators of transit performance. In this chapter, we use the outputs of that tool, combined with several complementary data sources, to estimate time series models of transit ridership. These models are designed to provide insight into what is driving changes in ridership, both on MUNI and on the Bay Area Rapid Transit (BART) system. In Chapter 5, they are applied to understand why, over the 2009 to 2013 period, BART ridership experiences robust growth, while MUNI ridership does not, given that both operate in the same metropolitan area. In Chapter 6, they are applied to three ex-post evaluations of system changes and two forecast applications to further demonstrate how they can be used to inform transport planning.

4.1 Introduction

This research fits in the context of a broader body of literature examining the drivers of transit demand.

In their 2003 working paper, Taylor and Fink [170] review the research on the factors affecting transit ridership. They divide the research into descriptive analyses and causal analyses. Descriptive analyses are generally based on survey and interview data, and tend to be more qualitative in nature. Causal analyses generally involve multivariate regressions, and are of greater interest here, due to their ability to quantify the contributing factors. Examples include [171, 172, 173, 174].

Taylor and Fink go on to list the factors commonly found to influence transit ridership, with the key factors found to be:

- Access to private autos;
- Employment and Central Business District (CBD) employment;
- Income and auto ownership;
- Price of gasoline;
- Parking cost and parking availability;
- Housing density and employment density;
- Public funding for transit subsidies;
- Fare and pricing policy;
- Service quantity, usually measured as revenue miles or vehicle miles; and
- Service quality, including bus information, safety, cleanliness and reliability.

Of the studies reviewed, Kain and Liu [174] provide a useful model by which econometric models can be applied to understand the major factors contributing to ridership changes. They estimated models of transit ridership as a function of four factors: employment, central city population, service miles and fares. They used the model to calculate the elasticities associated with each of those factors, and applied those elasticities to the observed percent change in each of the factors for the Houston and San Diego transit systems. This research applies estimated models in a similar manner, as described in Section 5.4.

Taylor and Fink point out that a key limitation of these studies is that their criteria for which data are included tends to be those data that are readily available, particularly as it relates to service quality. While this thesis does not claim to fully move beyond limitation of data availability, it does involve a very extensive data assembly effort aimed at making a broad set of relevant variables available to the analysis. This is facilitated by the inclusion of several data sets that were not available at the time of their review. These include Automated Vehicle Location (AVL) and Automated Passenger Counter (APC) data which were relatively new at the time and provide a means of measuring transit crowding and reliability, the General Transit Feed Specification (GTFS) as a detailed measure of transit schedules which came into common use in 2009, the annual American Community Survey (ACS) which began in 2005, and the Longitudinal Employer-Household Dynamics (LEHD) data which provide geographically detailed estimates of employment and workers on an annual basis. Section 4.2 describes these data sources, and how they are combined from often different time scales and geographic resolutions.

There are several, more recent, contributions in this domain. Some focus on estimating the effect of a specific factor, and others are more broad. These studies provide a useful overview of the methodological toolkit available, and also an enumeration of the predictive variables commonly used in such models.

Brown and Neog [175] estimated regression models of the transit mode share in 82 Metropolitan Statistical Areas (MSAs) in the U.S. using journey-to-work mode shares reported in the 2000 Census. They find significant relationships between transit ridership and four descriptive variables: service frequency (ratio of vehicle kilometres to route kilometres), service coverage (ratio of route kilometres to population, percent of households that do not own an auto and unemployment rate. Taylor et al. [176] estimated models of transit ridership in 256 U.S. urbanised areas using 2000 ridership and Census data. The find that the significant factors influencing transit ridership per capita in metropolitan areas are: vehicle revenue hours, geographic land area, median household income, non-transit non-single occupant vehicle trips, transit fares and headways/service frequency. Both of these studies are cross-sectional analyses that provide insight into the differences between cities, but not into the trends over time.

A number of studies focus on the effect of fuel prices on transit ridership. Haire and Machemehl [177] used additive seasonal decomposition to remove seasonal trends from transit ridership, and then estimated multiple regression models to determine the effect of gasoline price on transit ridership. They found that seasonal patterns accounted for about 60% of the variability in transit ridership, highlighting the importance of accounting for seasonality in such analyses. The analysis accounts for vehicle revenue hours, the number of transit vehicles operating at maximum service, fuel cost, fare, Consumer Price Index (CPI), and the number of weekdays in the month, with the effect of fuel cost and CPI found to vary by region. Lane [178] uses time series regression models to estimate the effect of gasoline price on transit ridership, with consideration of lagged responses, while controlling for service changes and trend variables. Nowak and Savage [179] use Chicago Transit Authority (CTA) data to estimate regression models of the 12-month seasonal difference in the log of monthly transit ridership for the purpose of estimating the elasticity of transit ridership with respect to changes in gasoline price. The 12-month difference serves to remove the effect of seasonality (as discussed further in the methodology section of this chapter). Their use of a logarithm on both the dependent and descriptive variables allows the coefficients to be interpreted directly as elasticities. Their model includes gas price, average daily transit miles, fare, unemployment rate, and the proportion of weekdays in the month. Nowak and Savage also include a long list of other research that has examined the relationship between fuel price and transit ridership. These papers demonstrate the overall applicability of regression with time series data to estimate the relationship between ridership and other factors.

One more paper, which also focuses on fuel price is from Yanmaz-Tuzel and Ozbay [180]. They take a different approach, though, and rather than estimating a regression model on the time series data, they calculate the elasticity directly, based on the changes before and after specific events. This approach is somewhat more limited, because it does not consider the full extent of the data.

Not all topics receive as much attention as that of fuel price, but there are some other topics considered. Anderson [181] uses a regression model with a discontinuity to estimate the effect of a 2003 transit strike in Los Angeles. Tang and

Thakuriah [131] estimate the ridership effects of a real-time bus information system in Chicago. Their analysis is valuable because it uses longitudinal data (or panel data) rather than pure time series data, meaning that their data includes variation both across observations and through time. In their case, they treat the ridership on each bus route, for each month, as an observation. The real-time information system is rolled out at different times on different routes, which allows their estimated models to separate that effect from other factors with less risk of co-linearity. Brakewood et al [182] follow an equivalent approach to estimate the ridership effect of real time information on bus ridership in New York. One interesting outcome of that study is that the introduction of bike sharing in New York led to a reduction in bus ridership. This type of panel data analysis is viewed as a possible future enhancement for this research, when efforts to unify the spatial foundation for the estimation data are complete.

Chen et al. [129] combines a relatively broad consideration of factors affecting transit use, with time series models. They use auto-regressive fractionally integrated moving average (ARFIMA) models with independent regressors. Their final model considers gasoline price, transit fare, employment, service level and a set of seasonal constants.

An alternative approach is to examine travel trends more broadly. Metz [183] considers the demographic determinants of travel demand in Britain. LeVine and Jones [184] examine the trends in car and train travel in Britain. A specific focus is the observation that car travel appears to have levelled off in recent years, while train travel is growing rapidly. Their research draws from an analysis of different waves of the National Travel Survey (NTS), a household travel survey conducted annually in Britain. This allows for a detailed examination of trends for different sub-groups of the population. Unfortunately, annual travel surveys are not common in the United States, and not available for this study. The findings of these and related "broad trend" studies are discussed in further detail in Chapter 5; for now the focus is on identifying the appropriate method and the key variables to be captured in model estimation.

Collectively, these papers demonstrate the types of variables commonly in econometric models of transit demand. Typically, a handful of descriptive variables are used, including a measure of service quantity (often service miles), fare, gasoline price, and some measure of the economy (usually employment or the unemployment rate). The time series models tend to account for seasonality, either through seasonal differencing or by including seasonal constants. The two cross-sectional models discussed include a term on either the percent of zero-auto households or median household income, but the time series models do not. This may be because those factors tend to have greater variations between regions than over time, so are more explanatory in the cross-sectional domain. The analysis of strikes and service quality factors appears to be less common, and an area where further research could be of value. These variables provide an initial list of variables to be tested in model estimation.

This chapter seeks to build upon past research to understand why MUNI ridership stagnates and BART ridership grows over the same time period. Its modelling approach is most similar to that of Chen et al. [129], but it applies those models in manner similar to Kain and Liu [174]. It seeks to consider a broader range of descriptive factors than past research (specifically testing service quality factors, including reliability and crowding) and presents a visual means of understanding the effects of those factors through time. The findings with respect to the divergence raise interesting questions about whether current trends can be expected to continue in the future.

The remainder of this chapter is structured as follows. Section 4.2 describes the data considered in this analysis, and discusses notable trends in those data. Section 4.3 describes the methodology used to analyse the data. The methodology considers three types of time series models, and discusses the potential strengths and limitations of each. Section 4.4 examines the properties of the time series, and presents estimation results for each of the three types of models. The results are compared, and a preferred model of MUNI and BART ridership is selected. Finally, conclusions are presented of relevance to the model estimation and to the approach for analysing the drivers of transit demand.

4.2 Data

The data used in this study fall into four categories: MUNI data, data for BART and four other Bay Area transit modes, "drivers of demand" which are other factors expected to influence changes in the demand, and commute mode shares. The divergence of ridership and employment trends are explored towards the end of this section.

4.2.1 MUNI

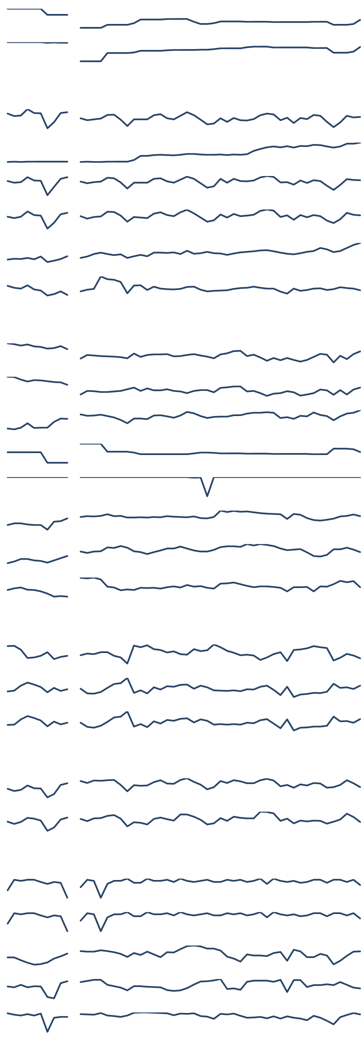
The data for the MUNI bus system comes primarily from transit AVL and APC data recorded on a subset of MUNI buses. The equipment is installed on approximately 25% of the bus fleet, which are assigned to different routes throughout the city on a daily basis to achieve coverage. Information on the complete schedule in operation is available in the form of the general transit feed specification (GTFS). The AVL/APC data are expanded to the GTFS data, such that they are representative of total system ridership on a daily basis, as described in Chapter 3.

These combined data are available for the period from June 2009 through November 2013, with data in April 2010 missing. For the purpose of calculating time series statistics and estimating models, the April 2010 data are imputed. The data are aggregated to average monthly conditions for weekdays, Saturdays and Sundays/holidays. Table 4.1 lists the key performance variables produced by the data fusion tool for average weekday conditions. Adjacent to each variable description is a sparkline showing the trend in the data over the period in question for average weekday conditions. The sparklines shown are for the daily totals, but the data are segmented into seven time periods: 3-6 am, 6-9 am, 9 am-2 pm, 2-4 pm, 4-7 pm, 7-10 pm and 10pm-3am. Each period is inclusive of the first minute listed and exclusive of the last. A full enumeration of the data by month is included in Appendix B.

Table 4.2 lists the same data fields, but with sparklines showing the change from 12 months earlier. Positive values show up in blue, and negative values are red,

Table 4.1: Trends from MUNI performance report

	Trend
Service Provided	
Vehicle Trips	
Service Miles	
Ridership	
Boardings	
Rear-Door Boardings	
Passenger Miles	
Passenger Hours	
Wheelchairs Served	
Bicycles Served	
Level-of-Service	
Average Run Speed (mph)	
Average Total Speed (mph)	
Average Dwell Time per Stop (min)	
Average Scheduled Headway (min)	
Average Full Fare ($)	
Average Distance Traveled per Passenger (mi)	
Average In-Vehicle Time per Passenger (min)	
Average Wait Time per Passenger (min)	
Reliability	
Percent of Vehicles Arriving On-Time (-1 to +5 min)	
Average Waiting Delay per Passenger (min)	
Average Arrival Delay per Passenger (min)	
Crowding	
Average Volume-Capacity Ratio	
Percent of Trips with V/C > 0.85	
Observations & Error	
Number of Days	
Days with Observations	
Percent of Trips Observed	
Measurement Error (ON/OFF-1)	
Weighting Error (SERVMILES/SERVMILES_S-1)	



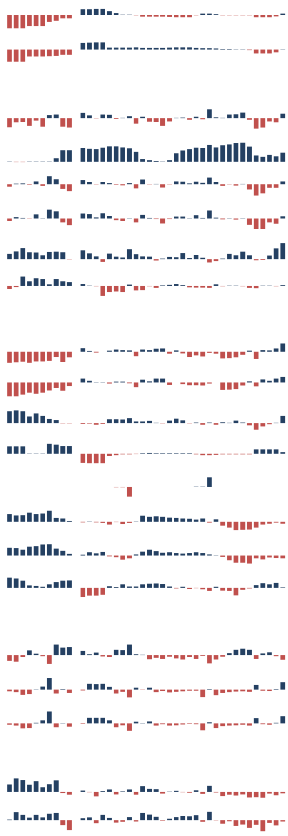
with the height indicating the change. This year-over-year change is used because it excludes the change due to seasonality. The sparklines listed here begin in June 2010 as the first month for which the the 12-month difference is available.

Several points are of note in these data.

Between March and May of 2010, the service provided, both in terms of vehicle trips and service miles, is reduced by approximately 10%. These cuts were made in response to a budget shortfall, and represent one of the largest service cuts in the history of MUNI [185]. That service is partially restored in September 2010,

Table 4.2: Annual differences from MUNI performance report

	Diff Trend (blue=pos, red=neg)
Service Provided	
Vehicle Trips	
Service Miles	
Ridership	
Boardings	
Rear-Door Boardings	
Passenger Miles	
Passenger Hours	
Wheelchairs Served	
Bicycles Served	
Level-of-Service	
Average Run Speed (mph)	
Average Total Speed (mph)	
Average Dwell Time per Stop (min)	
Average Scheduled Headway (min)	
Average Full Fare ($)	
Average Distance Traveled per Passenger (mi)	
Average In-Vehicle Time per Passenger (min)	
Average Wait Time per Passenger (min)	
Reliability	
Percent of Vehicles Arriving On-Time (-1 to +5 min)	
Average Waiting Delay per Passenger (min)	
Average Arrival Delay per Passenger (min)	
Crowding	
Average Volume-Capacity Ratio	
Percent of Trips with V/C > 0.85	



after which there are a series of incremental service increases until July 2013 when service miles are cut by approximately 2.5%.

With respect to ridership, the boardings, passenger miles and passenger hours are generally horizontal with bumps that may be a seasonal trend. The difference sparklines show three periods of general ridership decline and two periods of general increases. There is a general upward trend in the number of wheelchairs served, and a high point in the number of bicycles served that occurs between August and November of 2010.

One of the biggest changes observed in these data is the number of rear-door boardings, which increase from approximately 1,500 at the start of this period to over 200,000 by the end. MUNI began an all door boarding policy in July 2012, which was implemented in conjunction with a switch to a proof-of-payment system. This change was made with the goal of reducing the dwelling time required to board passengers, and MUNI's assessment of that policy found it successful at doing so [186]. Interestingly, the number of rear door boardings recorded by the APCs starts to ramp up in January 2011, and continues to increase throughout the analysis period.

Several performance measures are recorded as indicators of the level-of-service. The total speed is the average speed at which the bus travels, inclusive of stop time. The run speed is the average speed between doors closed and doors open. It excludes the time required for passengers to board and alight the bus, but still includes acceleration and deceleration time. The dwell time is measured as the time between doors open and doors closed each time the vehicle stops.

Following the April 2010 cuts, the dwell time increases and both speed measures decrease. This likely occurs because with fewer buses, there are more passengers per bus. Not only do they take longer to board and alight, but the vehicles are less able to skip stops when no passengers wish to board or alight. Changes to the scheduled headway are what would be expected given the service changes.

Fare changes are discussed in the following section.

There are some increases in the average distance per passenger and average in-vehicle time per passenger early in the analysis period. The increase in distance may be because travellers making short trips switch to walking when the bus frequency is reduced or fares are increased. The average wait time per passenger assumes that passengers wait for half the scheduled headway for the route they are boarding, at the time they board. The system-wide average wait time is less than half of the system-wide average scheduled headway because there tend to be more passengers on the more frequent routes. Essentially, it is weighted by the passengers boarding at each stop versus the vehicles stopping at each stop.

Reliability is reported both as a percent of vehicles arriving on-time and based on the average delay when passengers board (they are waiting longer for the vehicle to arrive) and when they alight (they get to their intended destination later).

Crowding is reported both as the average volume-capacity ratio, and as the percent of trips where the volume exceeds 85% of the capacity. This varies substantially by time-of-day and direction, as well as by route.

MUNI has struggled with both reliability and crowding issues in recent years [187]. Motivated in part by these issues, there is currently an effort underway to develop a person-based dynamic transit assignment model [188] in San Francisco that would provide a means to include reliability and crowding in planning-level models [189]. One of the interesting opportunities provided by these data is the ability to empirically test what, if any, affect these measures have on ridership.

The last set of measures contain measures of the potential errors in the data. The number of days and number of observations are self-explanatory.

The measurement error provides a means of considering errors in the values recorded by the APCs. Passengers are recorded when they board the vehicle and when they alight. The two values should add up to the same total for each trip a vehicle makes, but the data reflect some minor differences.

The weighting error provides a means of considering the sampling and weighting of the data, as described in Chapter 3. When the data are weighted and expanded, the weighting is proportional to the number of trip-stops in the scheduled versus sampled data. The number of service miles is also available both in the scheduled and sampled data, so if the weighting process were perfect, the number of service miles would match exactly. Instead, there are some minor differences, with 95% of observations falling between -2.2% and +0.8%.

4.2.2 Other Transit Modes

Table 4.3 shows several key trends for five transit modes serving San Francisco: MUNI bus, MUNI cable car, MUNI rail, BART and Caltrain. The bounds of the data shown is this table are from January 2001 through September 2015. Appendix B includes a full enumeration of the data by month.

MUNI bus is discussed above, and the mode for which the most detailed, automatically collected data are available.

As a matter of convenience, throughout this chapter, when “MUNI” is discussed without a qualifier, it refers to MUNI bus. While the cable cars are the iconic San Francisco transit mode, they carry only about 20,000 passengers on an average weekday, compared to about 500,000 on the buses. A substantial share of the cable car passengers are tourists, and the cash fare is set high, currently $6 per trip, to capture revenue from visitors.

MUNI also operates a light rail system, which is at the street level in mixed traffic throughout much of San Francisco. The light rail lines converge into a subway underneath Market Street in downtown San Francisco. This subway is directly aligned with, but on a different level from the BART subway in downtown San Francisco. Unfortunately, automated data collection is not available on the MUNI rail system, so counts are limited to manual counts no more than once per year.

BART, as discussed previously, is the regional rapid transit system, whose core market is carrying passengers from the East Bay and from the Peninsula into San Francisco.

The data reported in Table 4.3 come from several sources, which are available at different temporal and geographic resolutions. The columns indicate the sources and resolution of each data item. Table 4.4 shows the annual difference in each of these measures, with the remaining columns the same.

An important source of longer term data is the Statistical Summary of Bay Area Transit Operators [142]. This annual report is published by the Metropolitan Transportation Commission, and includes financial and operating data from each of the 24 Bay Area transit operators for the previous fiscal year. Key measures include ridership counts, service miles and service hours, operating costs, revenue and farebox recovery ratios. The advantage of these data are that they are available for all operators and offer a long time series, going back to the 1990s. They are, however, limited in that they are only available at the system level, and only summarised for each fiscal year. In Table 4.3, the statistical summaries provide a measure of the average monthly service miles and the average weekday ridership on each of the five modes. The data show service increases on BART and Caltrain for the first half of the periods, and variability or cuts in service on the three MUNI modes. The ridership growth on BART and Caltrain increases throughout this period, while the

Table 4.3: Trends for other transit modes

	Source	Temporal Res	Geog Res	Trend
Monthly Service Miles				
Muni Bus	Transit Stat Summ	FY	System	
Muni Cable Car	Transit Stat Summ	FY	System	
Muni Rail	Transit Stat Summ	FY	System	
BART	Transit Stat Summ	FY	System	
Caltrain	Transit Stat Summ	FY	System	
Average Weekday Ridership				
Muni Bus	Transit Stat Summ	FY	System	
Muni Cable Car	Transit Stat Summ	FY	System	
Muni Rail	Transit Stat Summ	FY	System	
BART	Transit Stat Summ	FY	System	
Caltrain	Transit Stat Summ	FY	System	
Average Weekday Ridership				
Muni Bus	APCs/Faregate	Monthly	Stop	
BART	APCs/Faregate	Monthly	Stop	
Cash Fare (2010$)				
Muni Bus+Rail	Published Values	Actual	System	
Muni Cable Car	Published Values	Actual	System	
BART	Published Values	Actual	System	
Average Fare (2010$)				
Muni Bus+Rail	Transit Stat Summ	FY/Actual	System	
Muni Cable Car	Transit Stat Summ	FY/Actual	System	
BART	Transit Stat Summ	FY/Actual	System	
Caltrain	Transit Stat Summ	FY/Actual	System	
Weekday Service Miles				
Muni Bus	GTFS	Actual	Route/Stop	
Muni Cable Car	GTFS	Actual	Route/Stop	
Muni Rail	GTFS	Actual	Route/Stop	
BART	GTFS	Actual	Route/Stop	
Weekday Service Miles-Extrapolated				
Muni Bus	Stat Summ/GTFS	Monthly	System	
Muni Cable Car	Stat Summ/GTFS	Monthly	System	
Muni Rail	Stat Summ/GTFS	Monthly	System	
BART	Stat Summ/GTFS	Monthly	System	

ridership on the three MUNI modes is variable.

More detailed automated count data is available on the MUNI bus and BART. The MUNI bus data comes from the APCs, and is processed as discussed above. These data are, however, only available for a limited duration. BART provides monthly “entry-exit” matrices that show the number of passengers traveling between each station pair for an average weekday, average Saturday, and average Sunday/holiday [143]. These data are derived from BART’s faregate system. BART fares are based on the distance travelled, so the system tracks the point of entry so it can deduct the correct fare when the traveller exits the system. Historically, this has been done using a magnetised paper ticket, but more recently BART has started accepting Clipper Cards, which allow a consistent mechanism of payment across different operators. Clipper Card transaction data was investigated as a contributing source of data to this study, and Chapter 2 discusses some of the limitations of

Table 4.4: Annual differences other transit modes

	Source	Temporal Res	Geog Res	Diff Trend (blue=pos, red=neg)
Monthly Service Miles				
Muni Bus	Transit Stat Summ	FY	System	
Muni Cable Car	Transit Stat Summ	FY	System	
Muni Rail	Transit Stat Summ	FY	System	
BART	Transit Stat Summ	FY	System	
Caltrain	Transit Stat Summ	FY	System	
Average Weekday Ridership				
Muni Bus	Transit Stat Summ	FY	System	
Muni Cable Car	Transit Stat Summ	FY	System	
Muni Rail	Transit Stat Summ	FY	System	
BART	Transit Stat Summ	FY	System	
Caltrain	Transit Stat Summ	FY	System	
Average Weekday Ridership				
Muni Bus	APCs/Faregate	Monthly	Stop	
BART	APCs/Faregate	Monthly	Stop	
Cash Fare (2010$)				
Muni Bus+Rail	Published Values	Actual	System	
Muni Cable Car	Published Values	Actual	System	
BART	Published Values	Actual	System	
Average Fare (2010$)				
Muni Bus+Rail	Transit Stat Summ	FY/Actual	System	
Muni Cable Car	Transit Stat Summ	FY/Actual	System	
BART	Transit Stat Summ	FY/Actual	System	
Caltrain	Transit Stat Summ	FY/Actual	System	
Weekday Service Miles				
Muni Bus	GTFS	Actual	Route/Stop	
Muni Cable Car	GTFS	Actual	Route/Stop	
Muni Rail	GTFS	Actual	Route/Stop	
BART	GTFS	Actual	Route/Stop	
Weekday Service Miles-Extrapolated				
Muni Bus	Stat Summ/GTFS	Monthly	System	
Muni Cable Car	Stat Summ/GTFS	Monthly	System	
Muni Rail	Stat Summ/GTFS	Monthly	System	
BART	Stat Summ/GTFS	Monthly	System	

those data. Where they are available, the APC and faregate data are preferred to the values reported in the statistical summaries due to their better resolutions.

The cash fares are assembled from the published fare values. The operators do not provide an archive or report of past fare values, but changes are publicly announced. Therefore, the past fares are identified by collecting press releases and newspaper articles announcing fare changes. All monetary values are inflation adjusted to year 2010 dollars, using the Consumer Price Index (CPI) for all urban consumers [107].

Notable fare increases on MUNI are from \$1.50 to \$2.00 on 1 July 2009, and an increase to \$2.25 on 1 September 2014. BART's recent policy is to implement regular fare increases in line with inflation. The BART fares reported are for the average fare paid, although the actual fare experienced by a traveller depends on the entry and exit station.

The average fare is measured as the revenue per boarding, and is in the range of \$0.80 per trip, in 2010 dollars, for most of this period. The difference reflects the usage of transit passes and discount fares.

Next is another measure of weekday service miles, this one derived from the GTFS data. The GTFS provides a more detailed measure of the transit schedule between each stop, and for the exact dates for which the schedule was in operation. The limitation is that GTFS is only available as far back as 2009. Therefore, an additional measure of extrapolated service miles is provided that pivots from the GTFS where it is available, and estimates the service prior to 2009 using the change in the reported values in the statistical summaries.

4.2.3 Expected Drivers of Demand

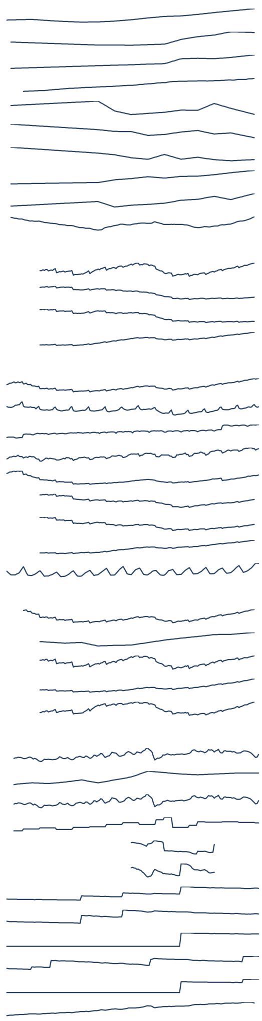
While it is recognised that the transport system can influence the land-use and socio-economic characteristics of a city, these factors are treated as exogenous for the purpose of this study. This study is limited to short-medium term effects, while the effect of transport on land use changes are assumed to occur over a longer time period.

Table 4.5 shows several key trends that may drive changes in demand. The bounds of the data shown is this table are from January 2000 through March 2015. Table 4.6 shows the annual difference in each of these measures, with the remaining columns the same. These data are for San Francisco, but all fields are available at a county level, and for the 4-county area as a whole. Appendix B includes a full enumeration of the data by month.

The tables start with measures of population and households. Most of these measures are taken from the annual ACS, for which these tables are available at the county level, starting in 2005. The 2000 Census provides the same tables measured for a larger sample, so 2001 through 2004 values are interpolated between the Census and the ACS. Two specific measures are available from different sources. The Census Bureau provides annual population estimates by county. Also, the San Francisco Planning Department provided data on housing completions. These were residential construction projects that were completed, with information for each

Table 4.5: Expected drivers of demand trends

	Source	Temporal Res	Geog Res	Trend
Population & Households				
Population	Census PopEst	Annual	County	
Households	ACS	Annual	County	
Housing Units	ACS	Annual	County	
Housing Units	Planning Dept/Census	Date	Block	
Households, Income $0-15k	ACS	Annual	County	
Households, Income $15-50k	ACS	Annual	County	
Households, Income $50-100k	ACS	Annual	County	
Households, Income $100k+	ACS	Annual	County	
Households, 0 Vehicles	ACS	Annual	County	
Median Household Income (2010$)	ACS	Annual	County	
Workers (at home location)				
Workers	LODES RAC/QCEW	Annual/Monthly	Block	
Workers, earning $0-15k	LODES RAC/QCEW	Annual/Monthly	Block	
Workers, earning $15-40k	LODES RAC/QCEW	Annual/Monthly	Block	
Workers, earning $40k+	LODES RAC/QCEW	Annual/Monthly	Block	
Employment (at work location)				
Total Employment	LODES WAC/QCEW	Monthly	Block	
Retail Employment	LODES WAC/QCEW	Monthly	Block	
Education and Health Employment	LODES WAC/QCEW	Monthly	Block	
Leisure Employment	LODES WAC/QCEW	Monthly	Block	
Other Employment	LODES WAC/QCEW	Monthly	Block	
Employees, earning $0-15k	LODES WAC/QCEW	Monthly	Block	
Employees, earning $15-40k	LODES WAC/QCEW	Monthly	Block	
Employees, earning $40k+	LODES WAC/QCEW	Monthly	Block	
Average monthly earnings (2010$)	QCEW	Monthly	County	
Jobs-Housing Balance				
Employees per Housing Unit	QCEW/Planning Dept	Monthly	Block	
Employees per Worker	LODES OD/QCEW	Annual/Monthly	Block	
Workers: Live & Work in SF	LODES OD/QCEW	Annual/Monthly	Block	
Workers: Live elswhere & work in SF	LODES OD/QCEW	Annual/Monthly	Block	
Workers: Live in SF & work elsewhere	LODES OD/QCEW	Annual/Monthly	Block	
Costs				
Average Fuel Price (2010$)	EIA	Monthly	MSA	
Average Fleet Efficiency (mpg)	BTS	Annual	US	
Average Fuel Cost (2010$ / mi)	BTS/EIA	Annual/Monthly	US/MSA	
Average Auto Operating Cost (2010$/mile)	IRS	Annual	US	
Median Daily CBD Parking Cost (2010$)	Colliers	Annual	CBD	
Median Monthly CBD Parking Cost (2010$)	Colliers	Annual	CBD	
Bay Bridge Toll, Peak (2010$)	BATA	Monthly	Bridge	
Bay Bridge Toll, Off-Peak (2010$)	BATA	Monthly	Bridge	
Bay Bridge Toll, Carpools (2010$)	BATA	Monthly	Bridge	
Golden Gate Bridge Toll, Peak (2010$)	BATA	Monthly	Bridge	
Golden Gate Bridge Toll, Carpools (2010$)	BATA	Monthly	Bridge	
Consumer Price Index	BLS	Monthly	US City Avg	

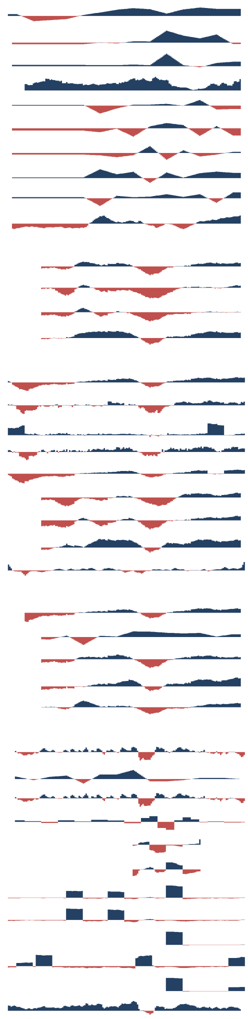


project that includes the address, the date the project was completed, and the net change in housing units. These data are used to provide a direct measure of the change in the number of housing units in San Francisco.

These data show that population and households are relatively flat at the start of the period, and then start to grow in the latter portion of the period. There are declines in the number of households with income less than $100,000 per year, and growth in the number of households with income of $100,000 or more per year. There is a general increase in the number of households that own zero vehicles. The inflation adjusted median household income has two periods of decline and two

Table 4.6: Expected drivers of demand annual differences

	Source	Temporal Res	Geog Res	Diff Trend (blue=pos, red=neg)
Population & Households				
Population	Census PopEst	Annual	County	
Households	ACS	Annual	County	
Housing Units	ACS	Annual	County	
Housing Units	Planning Dept/Census	Date	Block	
Households, Income $0-15k	ACS	Annual	County	
Households, Income $15-50k	ACS	Annual	County	
Households, Income $50-100k	ACS	Annual	County	
Households, Income $100k+	ACS	Annual	County	
Households, 0 Vehicles	ACS	Annual	County	
Median Household Income (2010$)	ACS	Annual	County	
Workers (at home location)				
Workers	LODES RAC/QCEW	Annual/Monthly	Block	
Workers, earning $0-15k	LODES RAC/QCEW	Annual/Monthly	Block	
Workers, earning $15-40k	LODES RAC/QCEW	Annual/Monthly	Block	
Workers, earning $40k+	LODES RAC/QCEW	Annual/Monthly	Block	
Employment (at work location)				
Total Employment	LODES WAC/QCEW	Monthly	Block	
Retail Employment	LODES WAC/QCEW	Monthly	Block	
Education and Health Employment	LODES WAC/QCEW	Monthly	Block	
Leisure Employment	LODES WAC/QCEW	Monthly	Block	
Other Employment	LODES WAC/QCEW	Monthly	Block	
Employees, earning $0-15k	LODES WAC/QCEW	Monthly	Block	
Employees, earning $15-40k	LODES WAC/QCEW	Monthly	Block	
Employees, earning $40k+	LODES WAC/QCEW	Monthly	Block	
Average monthly earnings (2010$)	QCEW	Monthly	County	
Jobs-Housing Balance				
Employees per Housing Unit	QCEW/Planning Dept	Monthly	Block	
Employees per Worker	LODES OD/QCEW	Annual/Monthly	Block	
Workers: Live & Work in SF	LODES OD/QCEW	Annual/Monthly	Block	
Workers: Live elswhere & work in SF	LODES OD/QCEW	Annual/Monthly	Block	
Workers: Live in SF & work elsewhere	LODES OD/QCEW	Annual/Monthly	Block	
Costs				
Average Fuel Price (2010$)	EIA	Monthly	MSA	
Average Fleet Efficiency (mpg)	BTS	Annual	US	
Average Fuel Cost (2010$ / mi)	BTS/EIA	Annual/Monthly	US/MSA	
Average Auto Operating Cost (2010$/mile)	IRS	Annual	US	
Median Daily CBD Parking Cost (2010$)	Colliers	Annual	CBD	
Median Monthly CBD Parking Cost (2010$)	Colliers	Annual	CBD	
Bay Bridge Toll, Peak (2010$)	BATA	Monthly	Bridge	
Bay Bridge Toll, Off-Peak (2010$)	BATA	Monthly	Bridge	
Bay Bridge Toll, Carpools (2010$)	BATA	Monthly	Bridge	
Golden Gate Bridge Toll, Peak (2010$)	BATA	Monthly	Bridge	
Golden Gate Bridge Toll, Carpools (2010$)	BATA	Monthly	Bridge	
Consumer Price Index	BLS	Monthly	US City Avg	



periods of growth, corresponding to broader economic conditions.

The next set of measures is for workers. A worker is a person who is employed, and their location is recorded based on where they live, in contrast to an employee, for which the location is recorded at the place of work. Workers are reported both in total number, and segmented based on their annual earnings. The data show that the number of workers in San Francisco declines during the two recessions, and increases between the recessions and in the current recovery period.

Estimates of workers come from the LEHD Origin-Destination Employment Statistics (LODES) data [101]. These data are provided by the Census Bureau, and are currently available annually from 2002 through 2014. LODES provides an estimate of employed people, summarised based on where they live and where they work. There are three compilations of LODES data: Residence Area Characteristics (RAC), Workplace Area Characteristics (WAC), and Origin-Destination (OD). In all cases, the data are segmented by worker earnings and industry. They are available at the scale of Census blocks, so for a very detailed geographic resolution. The limitation of these data are that they are effectively a synthetic data set. The Census Bureau combines several data sources to derive this information, including reported unemployment insurance information from firms, and tax records. To protect privacy, the underlying source data are not released, and the public data are "fuzzied" such that they match in aggregate, but some variation at small scales can be expected. Nonetheless, they are widely available, updated annually, and appear to be reasonable based on initial inspections.

For this research, the LODES is combined with another Census data product, the Quarterly Census of Employment and Wages (QCEW) [190]. The QCEW provides estimates of employment by industry at the county level, as well as average wages. It is derived from unemployment insurance information, and thus can have some limitations for industries that are not required to file unemployment insurance, such as certain categories of agricultural and military employment. The QCEW is released on a quarterly basis, but the quarterly release includes estimates of employment for each month. Thus, it provides a less geographically detailed, but more temporally detailed measure of employment than the LODES.

In the trends shown, the annual estimates of workers from the LODES are adjusted to match the monthly distributions of employment in the QCEW for each county.

Next, the tables show employment by industry and by worker earnings, as well as the average monthly employee earnings. All are reported based on the work location, and are derived from the combined LODES/QCEW.

The employment data show a few trends of note. First, the total employment declines after the dot-com bust and after the financial crisis. The scale of the issue is such that it took until 2014 to surpass the previous employment peak in 2001. Retail employment and leisure employment (hotels, restaurants, etc.) are both more seasonal than total employment. Education and health employment has two discrete increases, which are likely due to a change in how certain types of employment are categorised, rather than a real change. The trends in employment by earnings are equivalent to those in workers by earnings. Average monthly earnings are much more seasonal than the household income estimates, and the trend larger trend is somewhat less pronounced.

Next, a set of metrics tracks the jobs housing balance. The employees per housing unit is reported first. This has a similar shape to the employment curve, because employment tends to change more rapidly than residential units. The employees per worker in San Francisco grows over the latter portion of this time series, indicating that more employees are living in outlying counties. The next three metrics are derived from the LODES OD data: the number of workers who both live and work in San Francisco, the number who live elsewhere and work in San Francisco, and the number who live in San Francisco and work elsewhere. The latter peaks between 2005 and 2009, in spite of the more recent attention given to the Google buses (employee only buses operating between San Francisco locations and the company's Mountain View headquarters) [191].

The next section of the report shows trends in monetary costs. All costs are adjusted to 2010 dollars using the CPI for all urban consumers, as reported by the Bureau of Labor Statistics (BLS) [107].

The top several series are measures of the costs of auto travel. The average fuel price is reported on a monthly basis from the US Energy Information Administration (EIA) for the San Francisco Metropolitan Statistical Area (MSA). The fuel price generally grows over this period, with a notable decline in 2009.

The average fleet efficiency is measured for the US as a whole, and is reported by the Bureau of Transportation Statistics (BTS). The fleet efficiency increases most rapidly in the years prior to 2009, corresponding to the most rapid increase in fuel price. The average fuel cost, in dollars per mile, combines these two series.

An alternative measure of the auto operating cost is provided by the Internal Revenue Service (IRS). The IRS periodically provides an updated estimate of the marginal cost of operating a vehicle. This estimate determines how much people can deduct from their taxes when they use their vehicle for health, moving or charitable reasons. A separate measure, which is used when the vehicle is used for business purposes, includes the total cost of owning and operating the vehicle, but the ownership costs are excluded for the purpose of this research. The IRS rates are similar to the average fuel cost, but because they are only updated periodically, they reflect a less detailed set of changes.

From 2007 to 2012, a real estate analysis firm, Colliers International, conducted an annual survey of parking costs in the Central Business District (CBD) of major cities[192, 193, 194, 195, 196, 197]. Parking cost is expected to be a major determinant of transit mode share, so this is valuable information. Unfortunately, the survey was discontinued in 2012. Surveys were conducted prior to 2007, but they are not archived in an available location, and Colliers was not interested in providing them when contacted. It is expected that parking costs should be closely tied to employment conditions, but the time series are not long enough to draw many conclusions about their role.

The cost of tolls on the Bay Bridge and the Golden Gate bridge are included, as reported by the Bay Area Toll Authority (BATA). There are several toll increases over this period, and an introduction of tolls for carpools in 2010.

4.2.4 Commute Mode Shares

Table 4.7 shows the trends in commute mode shares, and Table 4.8 shows the annual differences. These data are reported in the annual ACS from 2005 through 2014. 2000 values are taken from the decennial Census, and 2001 through 2004 are interpolated. Values are reported as a total for all workers, and segmented by worker earnings. These trends are for San Francisco, but they are available for each county, and as a total. Appendix B includes a full enumeration of the data by month.

The data show a decline in auto modes and an increase in all other modes over this period. The trends are similar for lower and higher earning workers. Focusing

Table 4.7: Mode share trends

	Source	Temporal Res	Geog Res	Trend
Commute Mode Shares				
Drive-Alone	ACS	Annual	County	
Carpool	ACS	Annual	County	
Transit	ACS	Annual	County	
Walk	ACS	Annual	County	
Taxi, bike, other	ACS	Annual	County	
Work at home	ACS	Annual	County	
Commute Mode Shares by Segment				
Workers earning $0-50k: Drive-Alone	ACS	Annual	County	
Workers earning $0-50k: Carpool	ACS	Annual	County	
Workers earning $0-50k: Transit	ACS	Annual	County	
Workers earning $0-50k: Taxi, walk, bike, other	ACS	Annual	County	
Workers earning $0-50k: Work at home	ACS	Annual	County	
Workers earning $50k+: Drive-Alone	ACS	Annual	County	
Workers earning $50k+: Carpool	ACS	Annual	County	
Workers earning $50k+: Transit	ACS	Annual	County	
Workers earning $50k+: Taxi, walk, bike, other	ACS	Annual	County	
Workers earning $50k+: Work at home	ACS	Annual	County	
Workers with 0 vehicles: Drive-Alone	ACS	Annual	County	
Workers with 0 vehicles: Carpool	ACS	Annual	County	
Workers with 0 vehicles: Transit	ACS	Annual	County	
Workers with 0 vehicles: Taxi, walk, bike, other	ACS	Annual	County	
Workers with 0 vehicles: Work at home	ACS	Annual	County	

Table 4.8: Mode share annual differences

	Source	Temporal Res	Geog Res	Diff Trend (blue=pos, red=neg)
Commute Mode Shares				
Drive-Alone	ACS	Annual	County	
Carpool	ACS	Annual	County	
Transit	ACS	Annual	County	
Walk	ACS	Annual	County	
Taxi, bike, other	ACS	Annual	County	
Work at home	ACS	Annual	County	
Commute Mode Shares by Segment				
Workers earning $0-50k: Drive-Alone	ACS	Annual	County	
Workers earning $0-50k: Carpool	ACS	Annual	County	
Workers earning $0-50k: Transit	ACS	Annual	County	
Workers earning $0-50k: Taxi, walk, bike, other	ACS	Annual	County	
Workers earning $0-50k: Work at home	ACS	Annual	County	
Workers earning $50k+: Drive-Alone	ACS	Annual	County	
Workers earning $50k+: Carpool	ACS	Annual	County	
Workers earning $50k+: Transit	ACS	Annual	County	
Workers earning $50k+: Taxi, walk, bike, other	ACS	Annual	County	
Workers earning $50k+: Work at home	ACS	Annual	County	
Workers with 0 vehicles: Drive-Alone	ACS	Annual	County	
Workers with 0 vehicles: Carpool	ACS	Annual	County	
Workers with 0 vehicles: Transit	ACS	Annual	County	
Workers with 0 vehicles: Taxi, walk, bike, other	ACS	Annual	County	
Workers with 0 vehicles: Work at home	ACS	Annual	County	

on the period from 2005 through 2013, where annual data are available, the total auto mode share (drive alone and carpool) declines from 48.0% to 41.3%. The transit mode share increases from 32.7% to 34.0%. The remaining modes start from a smaller share, but increase at a faster rate. Walk increases from 9.6% in 2005 to 11.2% in 2013. Taxi, bike and other increase from 3.4% to 6.5% and work at home increases from 6.3% to 7.0%. These data do not explain why the mode shift occurs, but it is notable nonetheless.

4.2.5 Divergence of Ridership and Employment

Past research [170] indicates that employment changes can be expected to be one of the most important drivers of changes in transit ridership. However, as Figure 4.1 shows, MUNI ridership decreases relative to employment while BART ridership increases relative to employment during the period of analysis. These divergent relationships and their implications for model estimation are examined in further detail here.

Figure 4.2 shows the MUNI ridership, the employment in San Francisco, and the MUNI riders per employee for the period over which detailed data are available, from June 2009 though November 2013. These data show steady employment growth in San Francisco from about 2011 onwards, but no accompanying ridership growth. Instead, there is a general downward trend in MUNI riders per employee.

This trend is more clear when the MUNI riders per employee series is decomposed into its trend, seasonal and random components, as shown in Figure 4.3. The decomposition is performed using the standard additive method [198]. The trend line shows a reduction from about 0.94 riders per employee to 0.82 riders per employee. That is a reduction of about 13% in four years.

Figure 4.4 shows the BART riders, the employment in the 4-county area served by BART, and the BART riders per employee. The BART data and accompanying employment data are available for a much longer period, from January 2001 through March 2015. Employment reached its peak in December 2000 with the dot-com boom, an especially pronounced event in the Bay Area. Employment declined until 2004, and started to rebound in 2005 before falling again with the recession in late

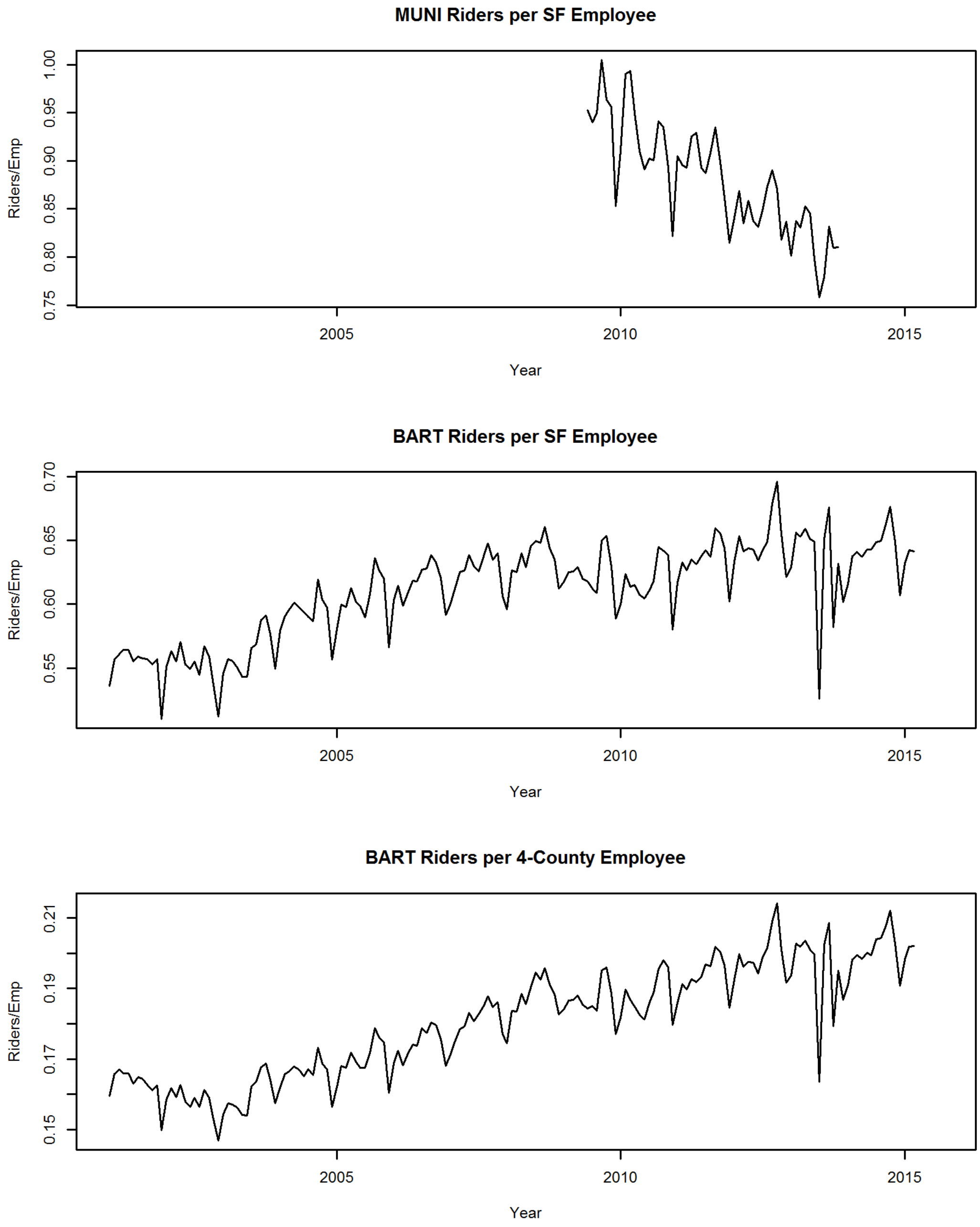


Figure 4.1: MUNI and BART riders per employee

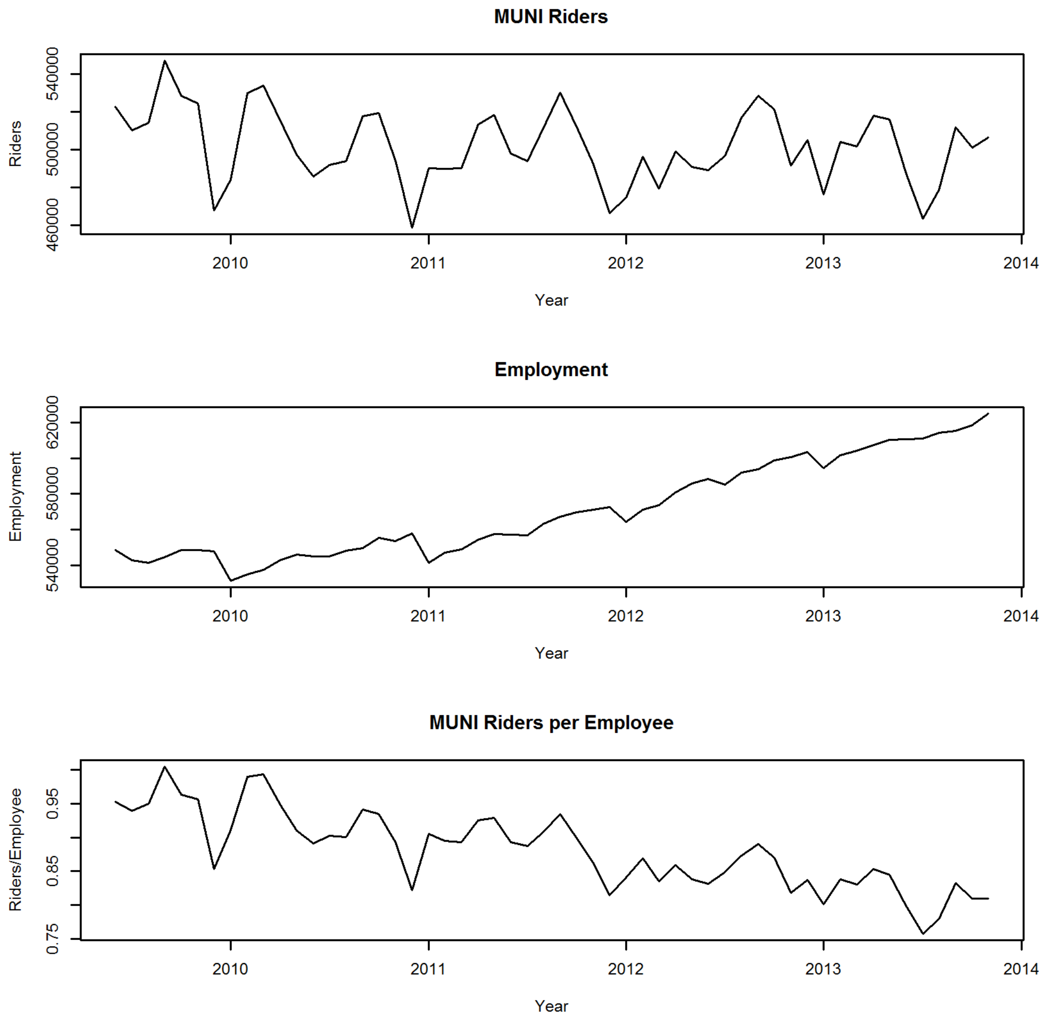


Figure 4.2: Trends in MUNI ridership and San Francisco employment

2008 and 2009. Employment growth has been strong from 2010 onwards, finally surpassing its 2000 peak in late 2014. Similar periods of growth and decline can be observed in the BART ridership data. The third sub-graph shows a general increase in BART riders per employee over this period.

The pattern is more clear when the BART riders per employee time series is decomposed into its trend, seasonal and random components, as shown in Figure 4.5. The large spikes in the random component in 2013 are due to BART employee strikes in July and October of that year. Over the period from 2001 through 2015, the BART ridership increases from about 0.16 per employee to about 0.20 per employee, or about 25%.

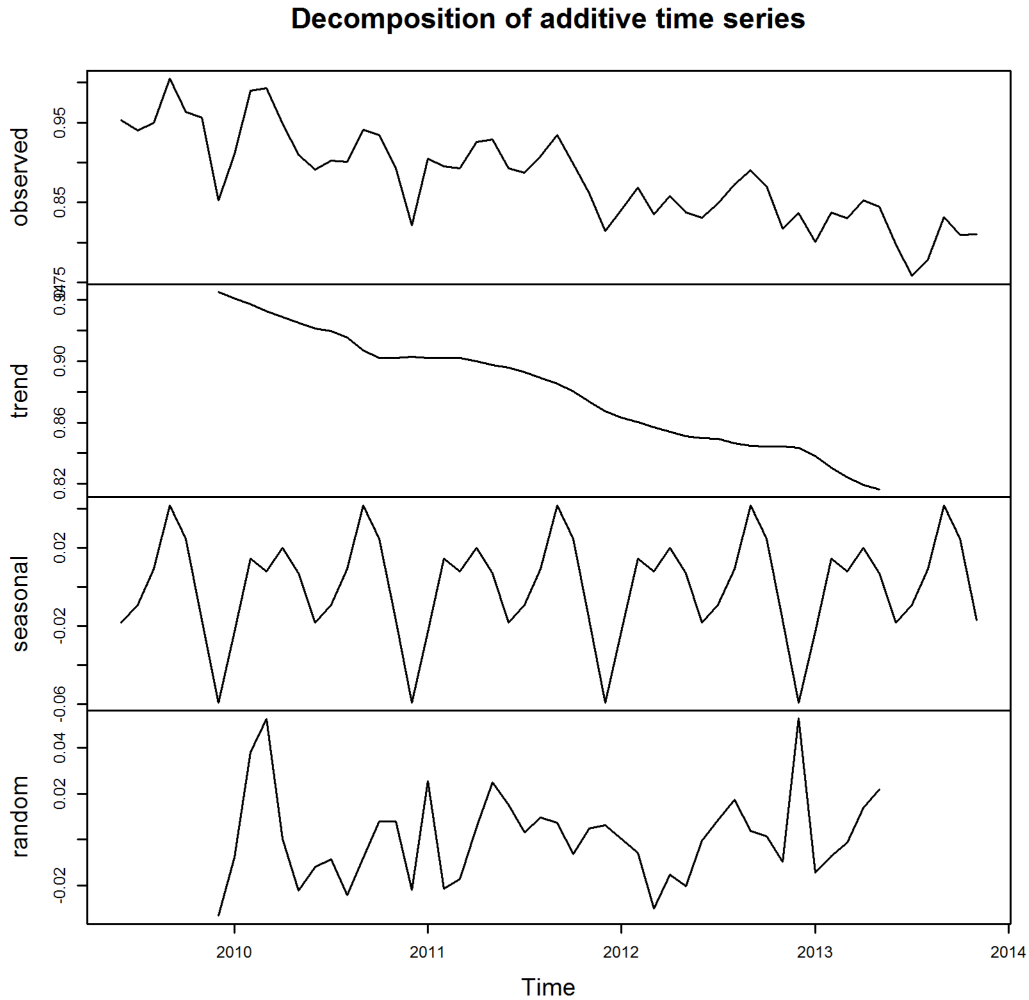


Figure 4.3: Decomposition of MUNI riders per employee

The models estimated in the subsequent sections seek to determine the relative influence of different factors, including employment, on the transit ridership changes. They seek to answer the question of why the ridership trends diverge so starkly.

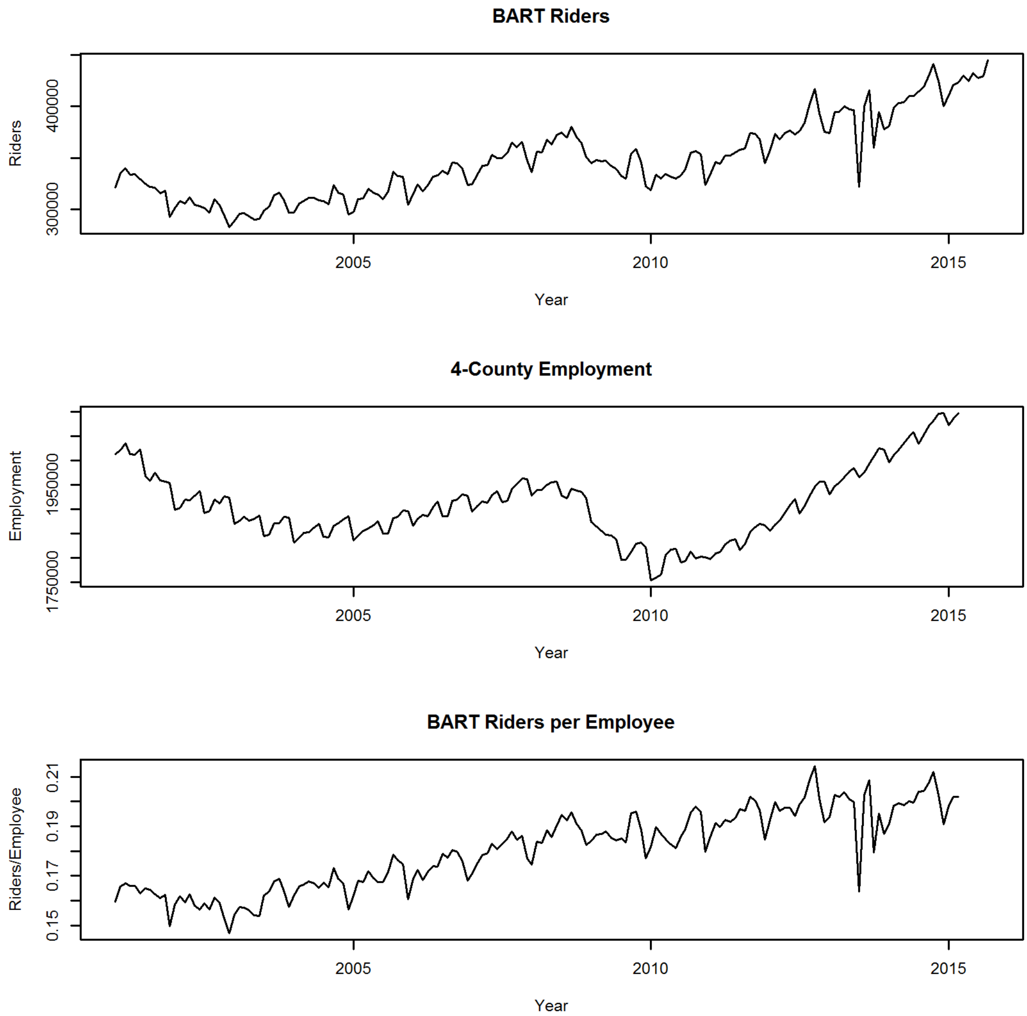


Figure 4.4: Trends in BART ridership and 4-county employment

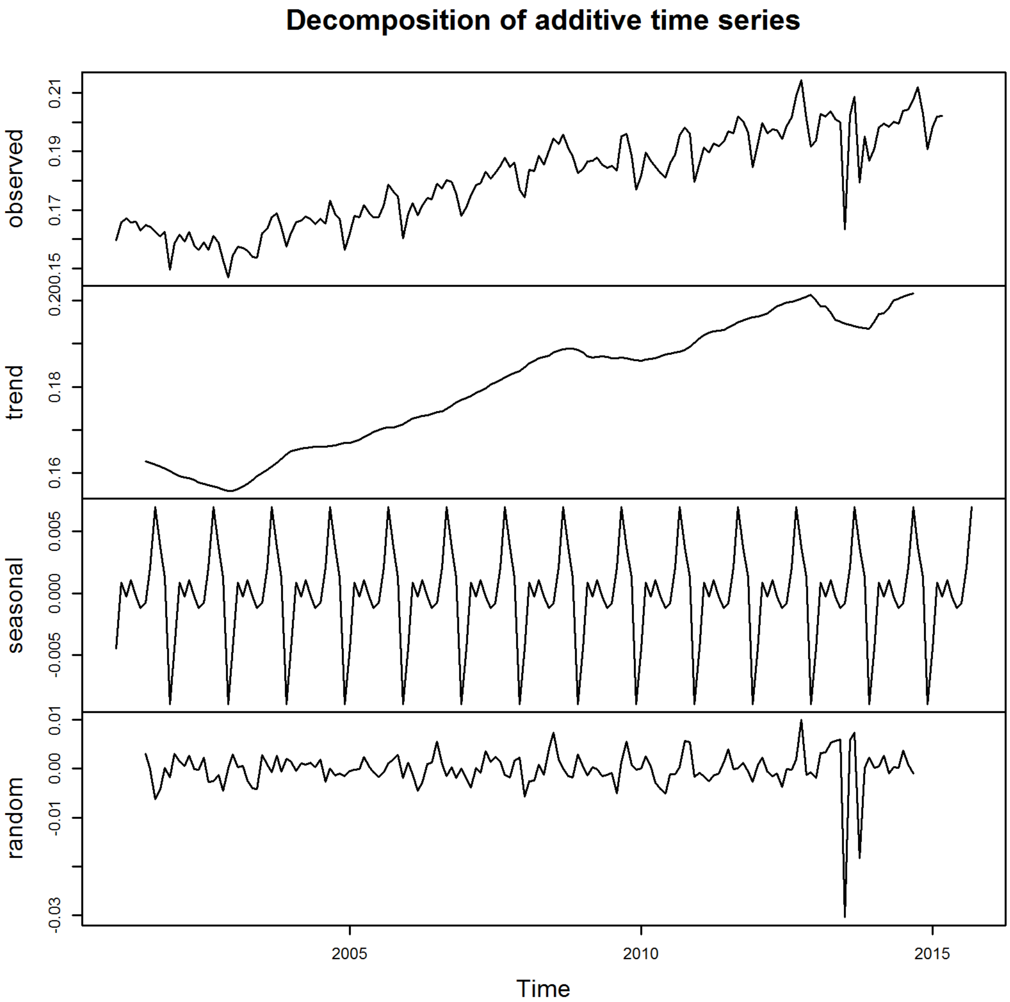


Figure 4.5: Decomposition of BART riders per employee

4.3 Methodology

In order to quantify the contributions of a range of factors on the changes in MUNI and BART ridership, a series of statistical models is estimated. The total ridership on each system is modelled as a time series: a one-dimensional observation through time. Future work will re-examine these data using a panel data structure where the ridership at each stop or station in each period is treated as a separate observation. That re-analysis is planned as a future student project, but is beyond the scope of this research.

Three types of models are considered in this chapter: Autoregressive Integrated Moving Average (ARIMA) models, regression on time series data, and Regression with ARIMA Errors (RegARIMA).

ARIMA models are aimed at predicting future values of the time series as a function of past values. The ARIMA approach provides a means of identifying and estimating current and future values of the time series using past values.

The second approach is to use regression models on time series data. Regression models can include previous values of the time series, which would result in a similar effect to ARIMA models, although ARIMA models are generally more sophisticated in their handling of past values. Importantly, regression models can also include other descriptive variables in cases where such measures are available. This provides two advantages. First, it allows for those variables to be used in forecasting. Second, it allows for statistical inference with respect to the correlation of the time series with those other descriptive variables.

Regression models with ARIMA errors combine elements of both types of models. They are considered here as a means for overcoming a potential violation of the standard regression assumptions, specifically that the model's errors should be independent of time.

The remainder of this section introduces these three types of models, as well as associated strategies for working with time series data.

4.3.1 Measuring Autocorrelation

An important attribute of time series data is that current values of the time series tend to be correlated with past values of the series, meaning that they are autocorrelated. There are two important implications of this. First, it means that past values can be an effective predictor of current or future values. Second, it means that the data contain less information than if each were an independent observation.

The autocorrelation for lag k can be calculated as [198]:

$$r_k = \frac{\sum_{t=k+1}^{T}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2} \tag{4.1}$$

where y_t is the value of the time series at time t, $\bar{y}$ is the mean value of the time series, and T is the length of the time series.

Similarly, the partial autocorrelation measures the correlation between the current time series value and a lagged value, excluding the correlation contributed by intermediate values. To give an example, consider a case where the value at each period y_t is correlated with the value at the previous period y_{t-1}. It follows that y_{t-1} is correlated with y_{t-2}, and that y_t is also correlated with y_{t-2} because y_{t-1} falls between and is correlated with both. The partial autocorrelation excludes this intermediate effect and measures only the additional, direct, correlation between y_t and y_{t-2}.

The partial autocorrelations are estimated as the ϕ coefficients of an Autoregressive (AR) model for the appropriate number of lags. AR models are discussed in 4.3.3.

The Autocorrelation Function (ACF) is the series of autocorrelations for each value of lag, and the Partial Autocorrelation Function (PACF), is the series of partial autocorrelations for each value of lag. Plotting these provides a simple means of identifying which lags have strong autocorrelations. Figure 4.6 shows an example of a time series and its corresponding ACF and PACF plots. The dashed horizontal lines are plotted at $+/-1.96\sqrt{T}$. If the time series is white noise, we would expect 95% of values to fall within these bounds for both the ACF and the PACF.

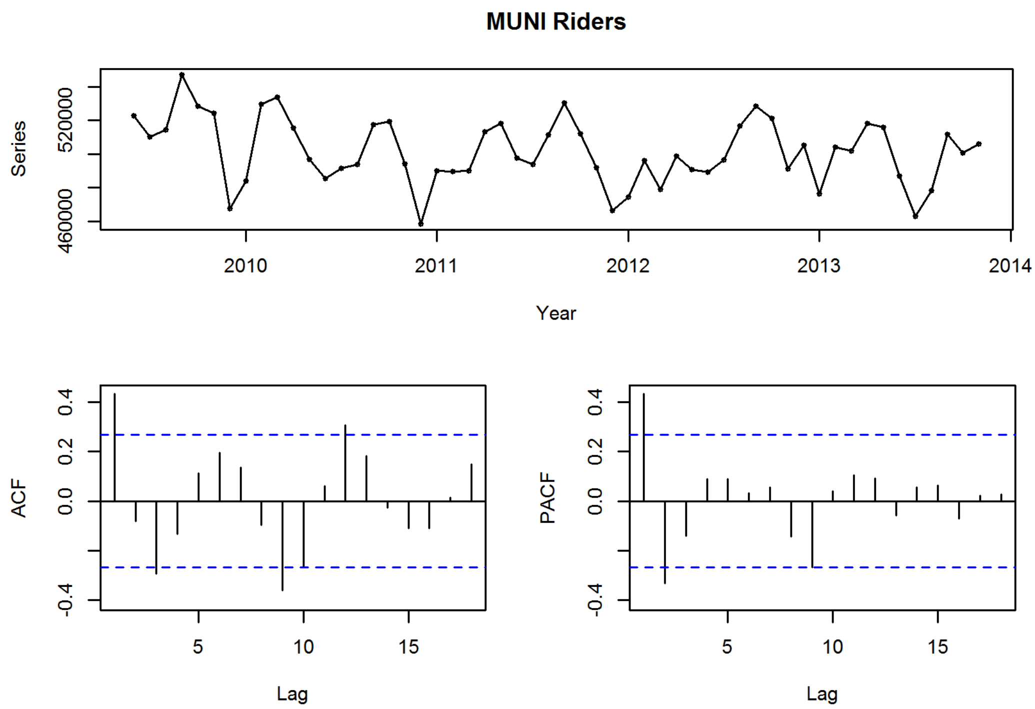


Figure 4.6: Example ACF and PACF plots

In addition to considering the autocorrelation of individual lags, the Box-Pierce test can be used as a measure of whether the autocorrelation in a group of lags is significantly different than what would be expected from white noise. The Box-Pierce test is written as [198]:

$$\begin{aligned} Q &= T\sum_{k=1}^{h} r_k^2 \\ Q &\sim \chi^2(h-K) \end{aligned} \tag{4.2}$$

where T is the number of observations, h is the maximum number of lags considered, r_k is the autocorrelation for lag k, K is the number of estimated parameters in the model, and Q follows a chi^2 distribution with the degrees of freedom equal to the number of observations minus the number of estimated model parameters.

The test combines values of r_k, such that large positive or negative values of r_k result in a large Q value. The value of h can be selected based on the number of seasonal periods in the data. These models are based on monthly data, so the seasonal, or periodic, effects are expected to occur every 12 months. For our analysis, we take h to be twice the number of seasonal periods (24) if T is large, and the number of seasonal periods (12) if T is small (less than five seasons). Given the chi^2 distribution, we consider p-values less than 0.05 to indicate significant autocorrelation of the group.

The Box-Pierce test has sometimes been observed to produce suspiciously low values of Q, so a modified test has also been proposed, known as the Ljung-Box test [199]. This test is expressed as:

$$\begin{aligned} Q^* &= T(T+2)\sum_{k=1}^{h}(T-k)^{-1}r_k^2 \\ Q^* &\sim \chi^2(h-K) \end{aligned} \tag{4.3}$$

where T is the number of observations, h is the maximum number of lags considered, r_k is the autocorrelation for lag k, K is the number of estimated parameters in the model, and Q^* follows a chi^2 distribution with the degrees of freedom equal to the number of observations minus the number of estimated model parameters.

To test the effect of the difference in these two tests, both were calculated for the preferred models reported in this document. The conclusions about whether or not the residuals are autocorrelated do not change regardless of which test is used. For example, the preferred BART RegARIMA model has a p-value for the Box-Pierce test of 0.207 and a p-value for the Ljung-Box test of 0.147, both of which are above our threshold of 0.05. To avoid duplication of similar values, only the Box-Pierce test results are reported in the tables in this chapter.

4.3.2 Stationarity and Differencing

A stationary time series is one whose properties do not depend on the time at which it is observed [198]. Stationarity can usually be determined by plotting and visually inspecting the time series. A stationary time series should be horizontal with constant variance. Seasonality also makes a time series non-stationary. ACF plots can be used to assist in identifying stationarity. For example, the time series in Figure 4.6 can be identified as non-stationary because 1) the series shows a downward trend, 2) the series shows apparent seasonality and 3) the ACF plot shows strong autocorrelation for the first several lags.

Stationarity is an important property, because estimating models on non-stationary time series can lead to spurious regression, as discussed in Section 4.3.4.

In this analysis, differencing is used to transform a non-stationary time series to a stationary one. Differencing is simply subtracting a lagged value in the time series from the current value. Differencing for a lag of one means that the differenced time series is the change from the previous point in time. It is possible to difference for lags greater than one, which can be used as a means for removing seasonality from the time series. For example, using a lag of 12 for monthly data means that the differenced time series represents the change from a year earlier.

In cases where the differenced data remain non-stationary, a second difference is taken. For the second difference, the lagged values of the differenced time series are subtracted from the current values of the differenced time series.

Once the data are shown to be stationary, regression can be safely used. It is only necessary to bear in mind what the resulting data represent, whether it is a difference or an absolute value, for the purpose of interpretation and analysis.

4.3.3 ARIMA Models

ARIMA models provide a means of predicting the current value of a time series as a function of its past values. Because ARIMA models do not, in themselves, provide a means for including descriptive variables in the models, they do not provide insight into why a time series changes in value. Nonetheless, they are presented here both as a reference model, and for use in combination with regression models, as discussed in Section 4.3.5.

The structure of an ARIMA model is expressed as:

$$ARIMA(p,d,q) \tag{4.4}$$

where p is the order of the autoregressive component, d is the degree of differencing, and q is the order of the moving average component.

The model itself is written as:

$$(1-\phi_1 B-...-\phi_p B^p)(1-B)^d y_t = c+(1+\theta_1 B+...+\theta_q B^q)e_t \tag{4.5}$$

where y_t is the value of the time series at time t, B is the backshift operator, such that $By_t = y_{t-1}$, $B^2 y_t = y_{t-2}$, and so forth, c is a constant, ϕ_1 through ϕ_p are the estimated autoregressive coefficients, θ_1 through θ_q are the estimated moving average coefficients, and e_t is the residual error at time t, with the e_t series assumed to be white noise.

A few points are to be made about this model.

The backshift operator follows the normal rules of algebraic operations. It is used here as a convenient notation to express more complicated ARIMA models.

$(1-B)^d$ is the differencing component of the model. d is taken as the degree of differencing required to make the time series stationary. A model estimated on data that is already stationary will be an ARIMA(p,0,q) model, which can also be expressed as ARMA(p,q), leaving out the integration.

$(1-\phi_1 B-...-\phi_p B^p)$ is the autoregressive component of the model. It is effectively a weighted average of past values of the time series.

$(1+\theta_1 B+...+\theta_q B^q)e_t$ is the moving average component of the model. It amounts to a weighted average of past model errors. The point of this term is to account for shocks to the system, beyond what was anticipated by the model. The expectation is that there may be some lingering effect of that shock that fades back to zero over the long term. This is in contrast to the autoregressive terms, where a sudden shock would carry forward to all future values of the time series. When forecasting the time series beyond q periods, the future values of e_t are assumed to be zero.

The interpretation of c depends on the degree of differencing. For $d = 0$, a constant represents the mean value. For $d = 1$, a constant means that the model incorporates a linear trend. For $d = 2$, a constant indicates a quadratic trend. Quadratic trends are excluded from this analysis due to their potential to lead to rapid changes.

The data considered in this study have a seasonal component. Therefore, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model is used. It is described as:

$$ARIMA(p,d,q)(P,D,Q)_m \tag{4.6}$$

where m is number of periods per season (such as 12 for monthly data), P is the order of the seasonal autoregressive component, D is the degree of seasonal differencing, Q is the order of the seasonal moving average component, and the remaining terms are as described in Equation 4.4.

The Seasonal ARIMA model is written as:

$$\begin{aligned}&(1-\phi_1 B-...-\phi_p B^p)(1-\Phi_1 B^m-...-\Phi_P B^{m+P})(1-B)^d(1-B^m)^D y_t = \\ &(1+\theta_1 B+...+\theta_q B^q)(1+\Theta_1 B^m+...+\Theta_Q B^{m+Q-1})e_t\end{aligned} \tag{4.7}$$

where Φ_1 through Φ_p are the estimated seasonal autoregressive coefficients, Θ_1 through Θ_q are the estimated seasonal moving average coefficients, and the remaining terms are as defined above. While it is possible to include a constant in this structure, the constant is not shown in this equation. In this structure, the seasonal terms are simply multiplied with the non-seasonal terms. An advantage to including a seasonal difference is that it avoids the need to estimate separate constants on each season.

This research follows the Hyndman-Khandakar algorithm [200] to determine the appropriate order of the ARIMA models, although with two deviations. First, the search is constrained to seasonal models with a second difference, which both forces stationarity and ensures that the models are consistent with the regression and RegARIMA models. Second, in cases where the algorithm would select a model with statistically insignificant coefficients, variations are tested to select a model with significant parameter estimates.

The Hyndman-Khandakar algorithm provides the method for selecting the appropriate order of the models: the values of (p,d,q) and (P,D,Q). It does this by selecting the values that minimise the AICc (defined in Section 4.3.7), within certain constraints. The AICc is itself based on a likelihood function. For each combination of (p,d,q) and (P,D,Q) considered, it is also necessary to estimate the coefficients ϕ_1 through ϕ_p, θ_1 through θ_q, Φ_1 through Φ_p and Θ_1 through Θ_q. This estimation is performed by maximising that same likelihood function, as applied to the differenced data.

The estimation is complicated by the fact that one set of coefficients is based on the model errors. The estimation of the ARMA model follows the algorithm of Gardner, Harvey and Phillips [201]. Because the coefficient estimation is performed on differenced data, it is reduced from an ARIMA model to an ARMA model. For simplicity's sake, we consider an ARMA(p,q) model, although can be extended as needed to seasonal models. The log-likelihood is maximised by minimising the function:

$$L^*(\phi,\theta) = n\log\big(\sum_{t=1}^{n}\tilde{v_t}^2\big) + \sum_{t=1}^{n}\log(f_t) \tag{4.8}$$

where n is the number of time steps, $\tilde{v_t}$ is the standardised residual at time t, and f_t is a value proportional to the prediction mean square error.

The reader is referred to Gardner, Harvey and Phillips [201] for the full details of the algorithm. For both this and the Hyndman-Khandakar algorithm, we use the implementation available in the forecast package in R.

4.3.4 Regression Models

The second type of model used in this study are linear regression models. The key advantage to using regression models over ARIMA models is that they provide a means of accounting for descriptive (independent) variables. The basic form of the regression model, as used here, is:

$$y_t = c + \beta X_t + e_t \tag{4.9}$$

where c is an estimated constant, y_t is the value of the time series at time t, X_t is a vector of regressors at time t, β is the estimated vector of parameters applied to those regressors, and e_t is the residual error at time t, assumed to be white noise.

When estimating regression models on time series data, Ordinary Least Squares (OLS) regression can be used, treating each time point as an observation. However, it is important that both the time series being modelled and all time series of descriptive variables used in the model be stationary. Estimating regression models on non-stationary data can lead to spurious regression, where the correlation between two variables is incorrect because both series contain a similar trend.

Hyndman and Athanasopoulos [198] provide an example of spurious regression in their textbook, which is shown in Figure 4.7. In this example, air passengers in Australia are plotted against rice production in Guinea. Both show high levels of growth between 1970 and 2010, and in the bottom plot appear to be highly correlated. A regression model would show a highly significant relationship between these two variables, even though they are clearly unrelated.

To be more precise, one of the assumptions of standard regression models is that the residuals are not autocorrelated [202], and models estimated from non-stationary data risk having autocorrelated residuals. The parameter estimates remain unbiased, but will be inefficient. This means that the standard errors and t-statistics will be inappropriate, which can lead to incorrect statistical inferences and possible specification errors. For these reasons, we check the residuals of our estimated regression models for autocorrelation using ACF and PACF plots.

It is worth noting that the regression models can also be estimated as a special case of RegARIMA models, with no autoregressive or moving average terms, and with the appropriate level of differencing. As a verification test, the preferred regression models were also estimated using this approach, and both methods produced equivalent results.

4.3.5 Regression Models with ARIMA Errors

A third type of model is considered in this study combines features of the regression model with ARIMA errors, or the RegARIMA model. Sometimes such models are

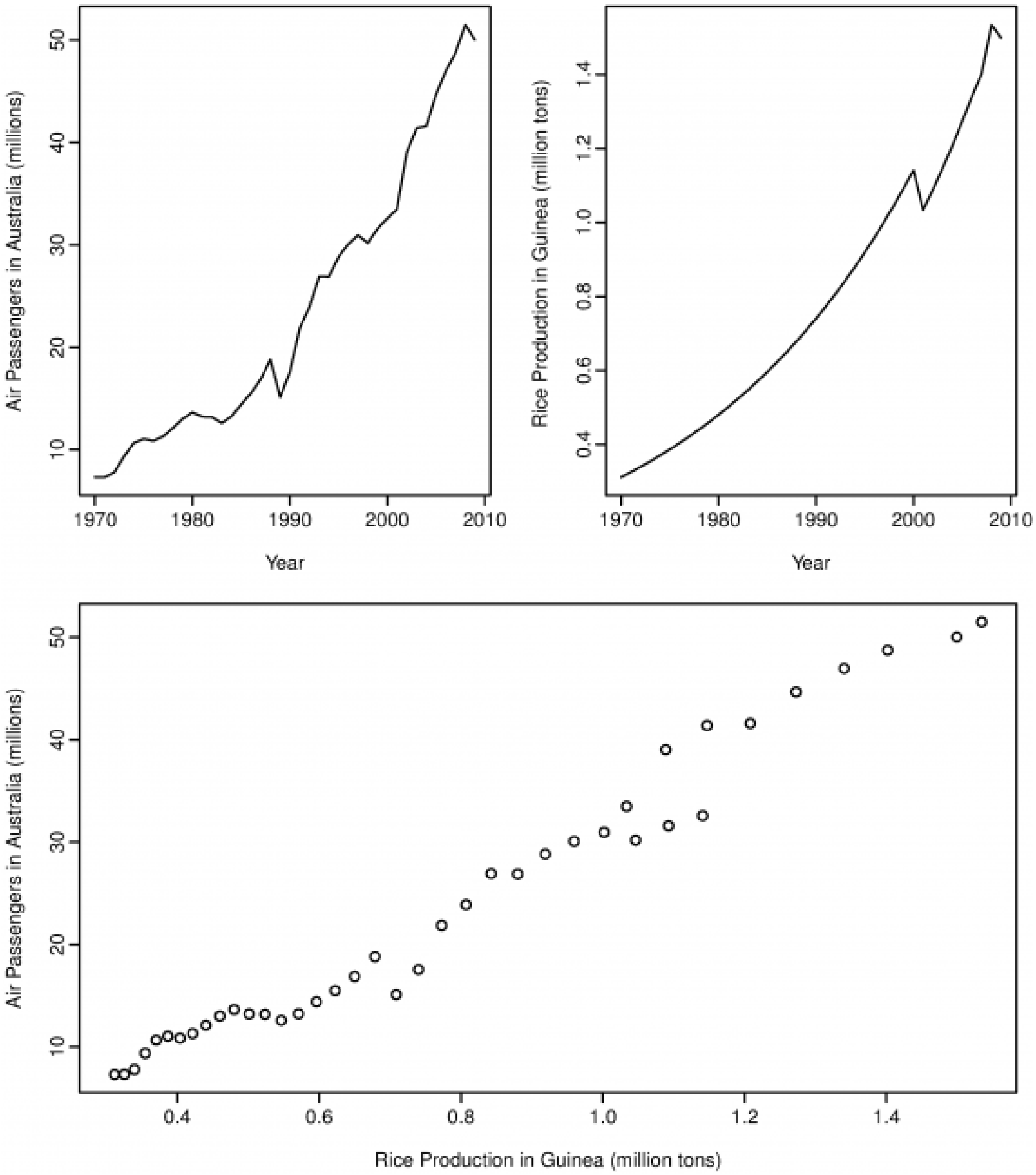


Figure 4.7: Example of spurious correlation [198]

referred to as dynamic regression. The advantage of the RegARIMA model is that it includes descriptive regressors, but can also capture the more subtle trends that ARIMA models can include. The latter is of particular importance if the residuals from a simple regression model are found to remain autocorrelated. In our case, we consider the regression model with seasonal ARIMA errors, which is expressed as:

$$\begin{aligned}
&y_t = \beta X_t + n_t \\
&(1 - \phi_1 B - \ldots - \phi_p B^p)(1 - \Phi_1 B^m - \ldots - \Phi_P B^{m+P})(1 - B)^d (1 - B^m)^D n_t = \\
&(1 + \theta_1 B + \ldots + \theta_q B^q)(1 + \Theta_1 B^m + \ldots + \Theta_Q B^{m+Q-1}) e_t
\end{aligned} \tag{4.10}$$

where n_t is the error of the regression component of the model, which is assumed to follow an ARIMA model as expressed in the second equation. e_t is the error of the model as a whole, and assumed to be white noise. The backcast operator B operates on n_t in he left hand side of the equation and on e_t in the right hand side. Note that Equation 4.10 is actually a pair of equations, where the first equation defines the model as a whole, and the second equation defines n_t specifically. This structure follows the definition included in Hyndman and Athanasopoulos [198], where the value of n_t is defined separately because it is a function of past values of itself.

When estimating a regression model with ARIMA errors, the ARIMA component affects the β estimates and the β estimates affect the ARIMA component. To avoid this problem, an iterative estimation procedure is used, as outlined in [198]. Before starting this iterative estimation, the degree of seasonal and non-seasonal differencing is selected manually such that the time series and the regressors are stationary. Underlying this iterative process are the Hyndman-Khandakar algorithm for order selection and the Gardner-Harvey-Phillips algorithm for coefficient estimation, as described in Section 4.3.3.

4.3.6 Lagged Variables

One advantage to estimating regression models on time series data is that it provides a convenient mechanism to test lagged effects of descriptive variables. This is valuable in transportation, where travellers are expected to have different long-term versus short-term responses to service or cost changes. This long versus short-term difference can take place in several ways.

First, it may take travellers some amount of time to learn about recent changes and settle on preferred means of fulfilling their desired activity patterns. While travellers can reasonably be expected to maximise their utility, it is also clear that people are creatures of habit who may continue in their current course, simply as a matter of default until prompted to make a change.

Second, some important travel decisions are made on a longer term basis than others. It is easy for a traveller to change routes in response to a service change, but workplace locations and levels of car ownership change less frequently. Travellers may also be locked into a certain transit fare pass or parking pass on a monthly or yearly basis.

In the first two cases, the long term effect is expected to be greater than the short term effect. It is also possible for a change to have a larger short term effect that degrades over time. Such a result may occur for a sudden price increase, where travellers respond quickly with a degree of “sticker shock”, and then gradually resume their previous behaviour once they become acclimated to the change.

While it is possible to estimate model coefficients for any combination of immediate response and lagged variables, experimentation on the data used in this project revealed that it is difficult to get meaningful relationships between the parameter estimates when multiple lags are considered. Therefore, a structure is imposed where the lagged effect is distributed linearly across a number of periods. The formula for this distributed lag is:

$$D_i(S,L) = \sum_{l=1}^{L-1} \left(\frac{\frac{L-l}{L}}{\frac{L}{2} - (S-0.5)} \right) x_{i-l} \tag{4.11}$$

where: $D_i(S,L)$ is the distributed lag result at time i, with a starting lag of S and an ending lag of L, and x_i is the value of the time series of interest at time i. For this analysis, S is constrained to be either zero or one.

To demonstrate the effect of this formulation, Table 4.9 shows the calculation for several values of S and L. By assuming that x is a series of ones, Table 4.9 shows the weight given to the value from each period. The weight is largest for the first month, and declines linearly thereafter. The weights always sum to one, which is convenient for the interpretation of the parameters. This means that for a coefficient estimated on a distributed effect, the long-term effect is that of the coefficient itself. For the example of $D_i(0,3)$, the immediate effect is half the coefficient's value.

Table 4.9: Example of distributed lag calculations

Lag (l)	x_{i-l}	$D_i(0,3)$	$D_i(0,6)$	$D_i(0,12)$	$D_i(1,3)$	$D_i(1,6)$	$D_i(1,12)$
0	1.0	0.500	0.286	0.154	0	0	0
1	1.0	0.333	0.238	0.141	0.667	0.333	0.167
2	1.0	0.167	0.190	0.128	0.333	0.267	0.152
3	1.0		0.143	0.115		0.200	0.136
4	1.0		0.095	0.103		0.133	0.121
5	1.0		0.048	0.090		0.067	0.106
6	1.0			0.077			0.091
7	1.0			0.064			0.076
8	1.0			0.051			0.061
9	1.0			0.038			0.045
10	1.0			0.026			0.030
11	1.0			0.013			0.015
12	1.0						
Total		1.000	1.000	1.000	1.000	1.000	1.000

For distributed lags with $S = 1$, there is no immediate effect. In estimation, models were tested that include both an immediate response variable, and a distributed lag on the same variable starting from $S = 1$. This is a more flexible structure that allows the estimation to determine the relative weight to allocate between the immediate response and the distributed lag. If the coefficients have the same sign, the long term effect is greater than the immediate response, and if the coefficients have opposite signs, the long term effect is less than the immediate response. The latter is a useful specification to test for "sticker shock" effects.

4.3.7 Goodness of Fit Measures

The regression models are estimated using ordinary least squares and use R^2 as the standard goodness-of-fit measure. It is defined as:

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \tag{4.12}$$

where $\hat{y}_i$ is the predicted value at time i, $\bar{y}$ is the mean of the observed values, and y_i is the observed value at time i. Higher R^2 values indicate a better fit. It is worth noting that differencing has an important effect on R^2 values. The mean value of the observed differences (the average rate of change) is very different from the mean value of the undifferenced time series. This means that R^2 values calculated from the undifferenced series will generally be higher than those calculated from a differenced series.

The ARIMA and RegARIMA model are estimated using Maximum Likelihood Estimation (MLE). The standard goodness-of-fit measures reported include Akaike's Information Criterion (AIC), and the Corrected Akaike's Information Criterion (AICc). They are defined as:

$$AIC = -2log(L) + 2(k) \tag{4.13}$$

$$AIC_c = AIC + \frac{2k(k+1)}{N-k-1} \tag{4.14}$$

where L is maximum likelihood value, k is the number of number of estimated parameters, and N is the number of observations,

Lower AIC and lower AICc values indicate a better fit. The key difference between the two arises for short time series, where AICc further penalises the inclusion of marginal parameters.

In addition to these measures, it is desirable to compare all the model types on a consistent basis. To do this, we compare the Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(\hat{y}_t - y_t)^2} \tag{4.15}$$

where y_t is the observed value at time t, $\hat{y}_t$ is the estimated value at time t, and N is the number of observations

To ensure that all models are evaluated against the same data, the RMSE is calculated based on fit against the training data set starting at period 14. This is the first period for which a 12-month seasonal difference and a 1-month second difference can be calculated internally from the data set. The RMSE is independent of the level of differencing. The percent RMSE is also reported with the denominator of the percentage based on the undifferenced time series.

4.4 Model Results

This section presents the model estimation results for three models of MUNI ridership and three models of BART ridership. For each transit system, an ARIMA model, a regression model and a RegARIMA model is tested. The underlying analysis involved estimating up to 100 models of each type to come to a preferred specification. Only the best model of each type is presented here. Section 4.4.5 compares the models of each type and recommends a preferred model for application. Prior to estimating the models, Section 4.4.1 examines the properties of the time series and the degree of differencing required to make those time series stationary. Appendix C shows the derivation of the formulas that can be used to apply each of the preferred models.

4.4.1 Stationarity and Differencing

In this section, three levels of differencing are examined for the MUNI riders and BART riders time series that will be modelled. The goal is to identify the level of differencing required to transform the series into stationary time series. This will be the degree of differencing used in the ARIMA models. Similarly, the regression models will be estimated on stationary time series to minimise the risk of spurious regression.

4.4.1.1 MUNI

Figure 4.8 shows the MUNI riders time series, a seasonal difference of the time series and the second difference of the time series. The second difference is calculated as the month-over-month change in the seasonal difference. The base time series shows apparent seasonality and a slight downward trend, so is not stationary. The seasonal difference seeks to remove both. It is difficult to determine visually whether the seasonally differenced time series is stationary, but the second difference does appear stationary, with a constant mean and random variation around that mean.

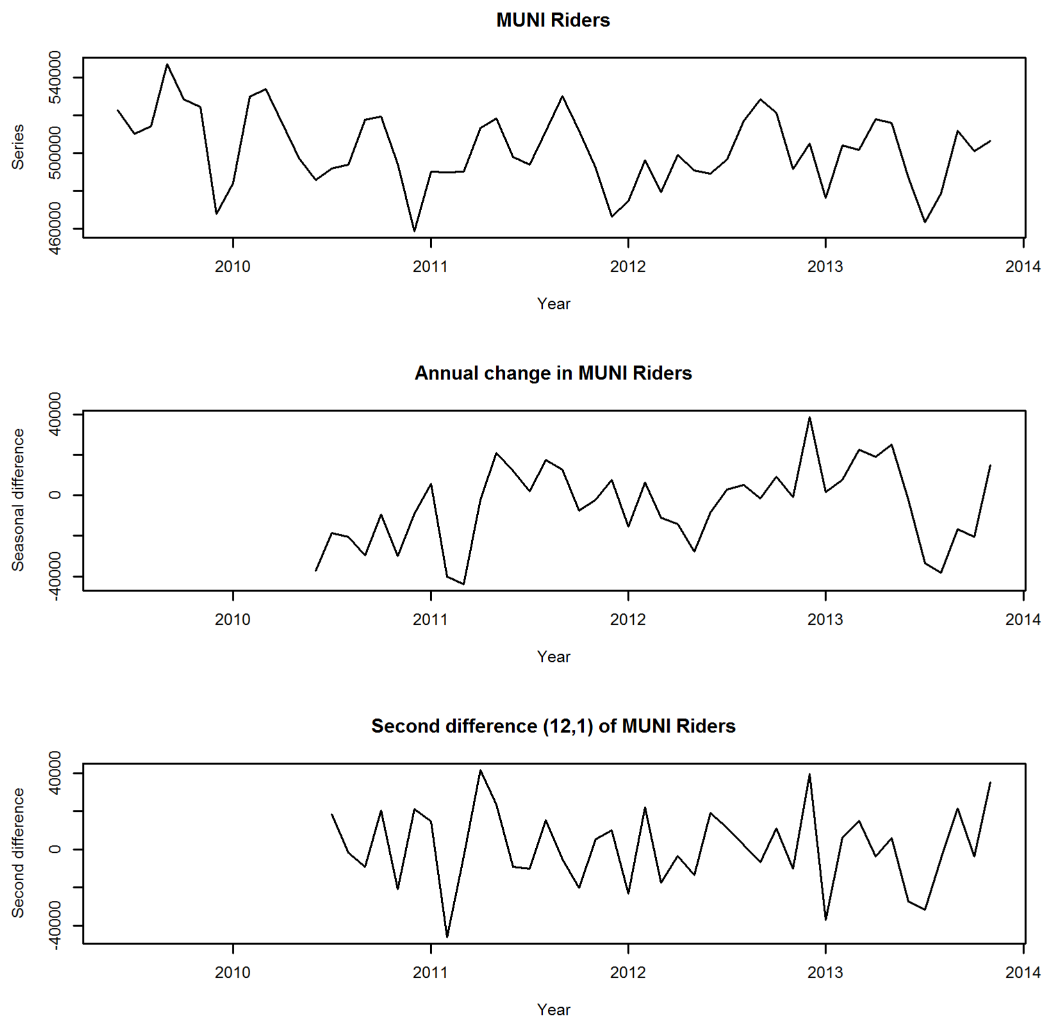


Figure 4.8: Differencing options for MUNI riders

To confirm our visual inspection, we also examine the ACF and PACF plots for each degree of differencing, as shown in Figure 4.9, Figure 4.10 and Figure 4.11. The base series (Figure 4.9), shows significant positive autocorrelation for lags of one and 12 months, and significant negative autocorrelations for lags of three and nine months, confirming our assessment that the data are not stationary. The seasonal difference (Figure 4.10) shows significant autocorrelation for the first lag, and the second difference (Figure 4.11) shows no significant autocorrelation.

While it may be debatable as to whether second differencing is necessary, for the regression models it is also important that all descriptive variables included in the model also be stationary. The evaluation of those series is not included here due to space limitations, but they revealed that a single seasonal difference was not stationary for several important variables, including population, workers and employment. Further, by imposing the same transformations on the dependent and descriptive variables, the relationships between those variables are preserved, aiding in the interpretation of the model results. Therefore, the second difference is selected as the basis for modelling MUNI riders going forward.

4.4.1.2 BART

Figure 4.12 shows the BART riders time series, a seasonal difference of the time series and the second difference of the time series. Visually, the second difference appears to be stationary, and the others do not.

The spikes toward the end of the time series are present because there are labour strikes on BART that took place in July and October of 2013. This resulted in a large, unexpected drop in ridership in those two months. When the time series is differenced, this shows up both as a decrease on those months, and as an increase in the subsequent months.

Figure 4.13, Figure 4.14 and Figure 4.15 show ACF and PACF plots for the three degrees of differencing in the BART series. The significant and slowly declining autocorrelations in the base series indicate that the data are clearly not stationary. Similarly, it takes 10 lags before the autocorrelations in the seasonal difference become insignificant, indicating that trends remain in the data. For the second dif-

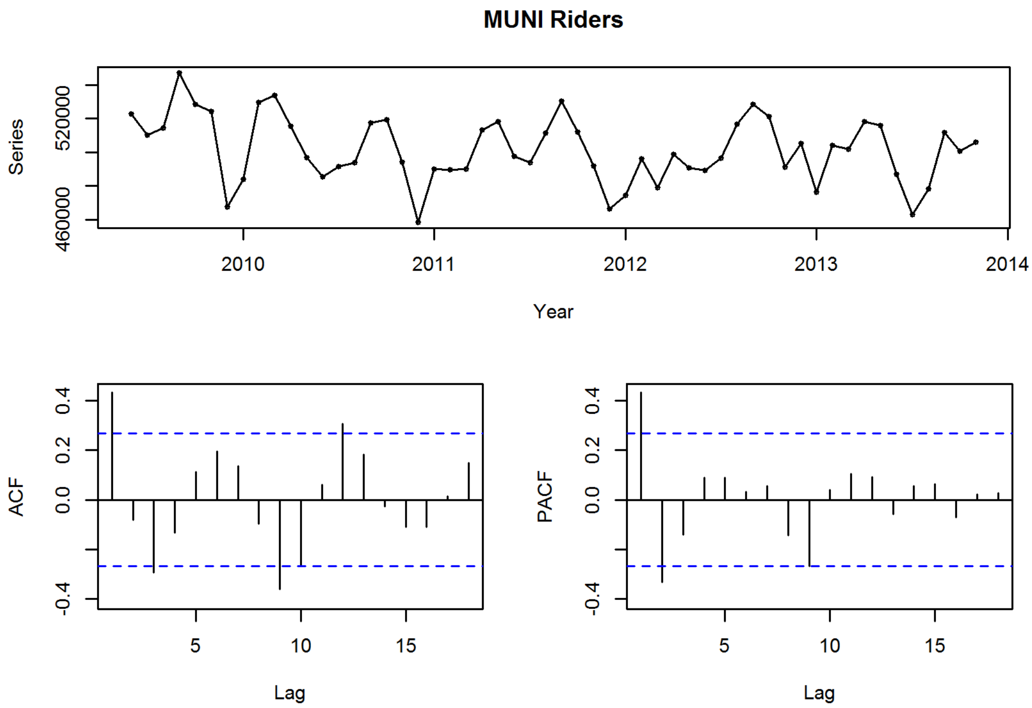


Figure 4.9: Autocorrelation of MUNI riders

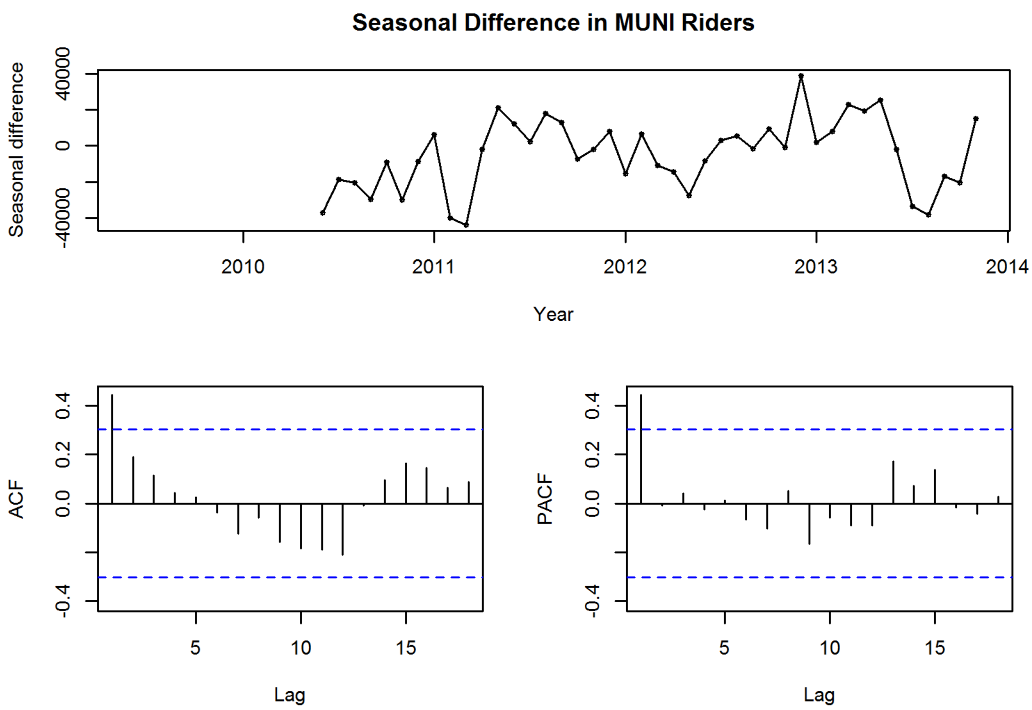


Figure 4.10: Autocorrelation of seasonal difference in MUNI riders

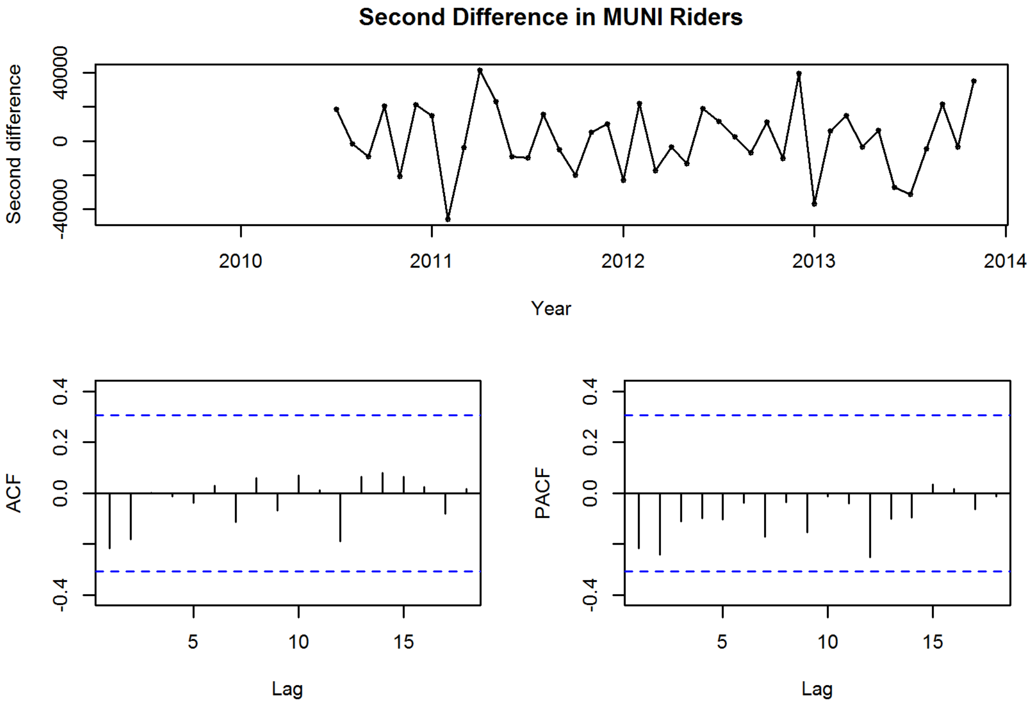


Figure 4.11: Autocorrelation of second difference in MUNI riders

ference, significant autocorrelations remain for some lags, but the pattern is more random, not the slow-declining trend observed in the base series or the seasonal difference.

A third difference is not considered, because doing so would result in a more complex model structure that that may include undesirable non-linear trends. Therefore, the second difference is selected as the basis for modelling BART riders as well, and we proceed to model estimation. The autocorrelation of the model residuals will be checked after the models are estimated, as a goal of enforcing stationarity at this point is to avoid correlated residuals at the back end. The worry is that correlated residuals could lead to spurious regression.

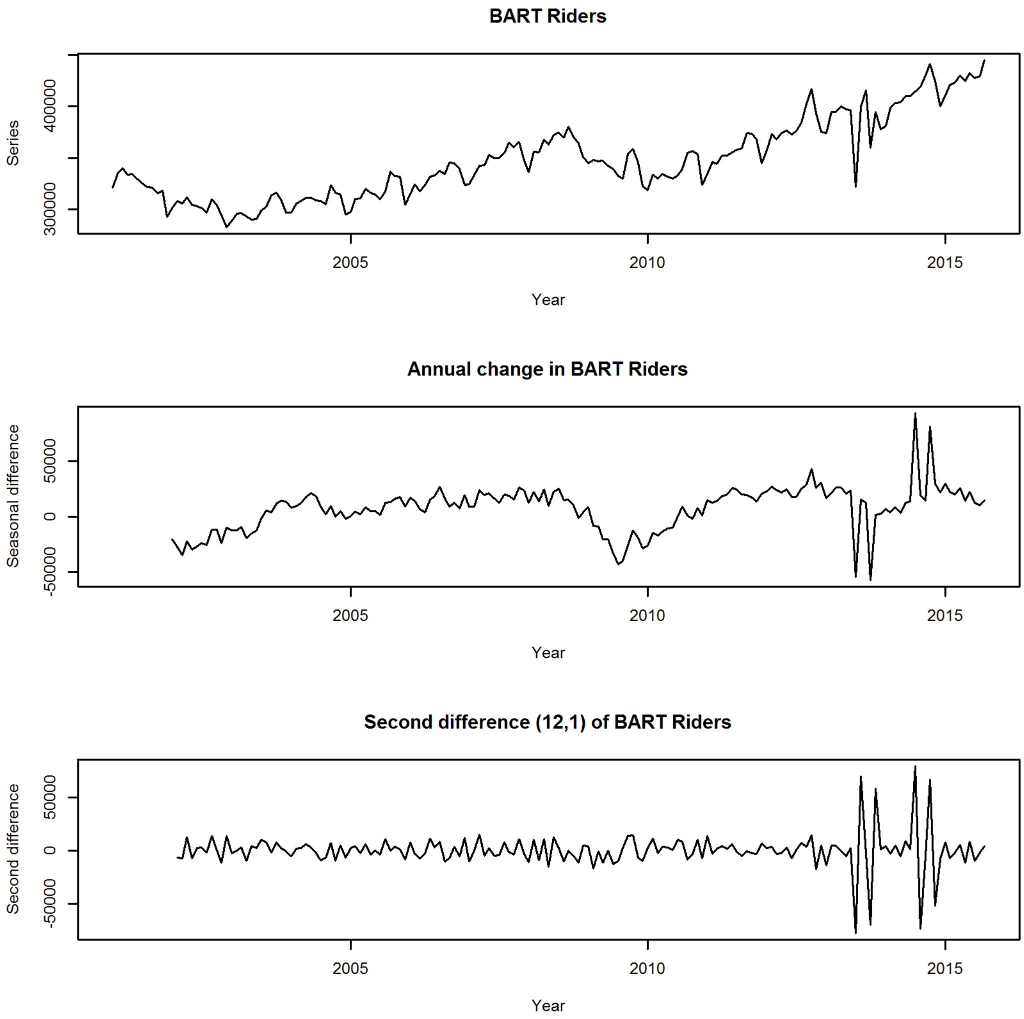


Figure 4.12: Differencing options for BART riders

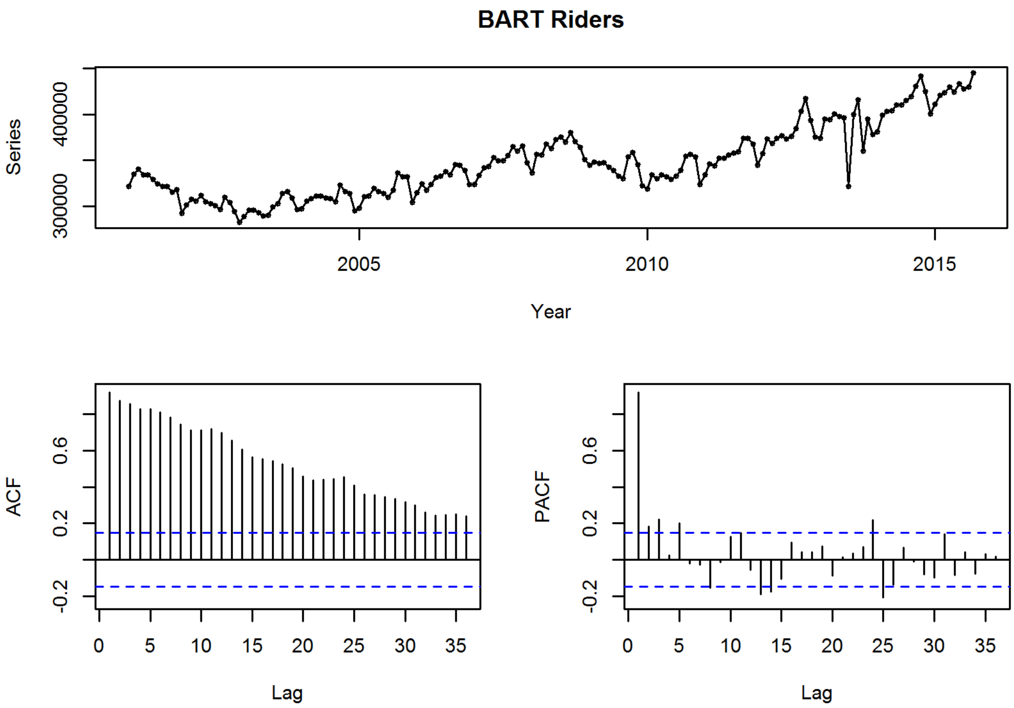


Figure 4.13: Autocorrelation of BART riders

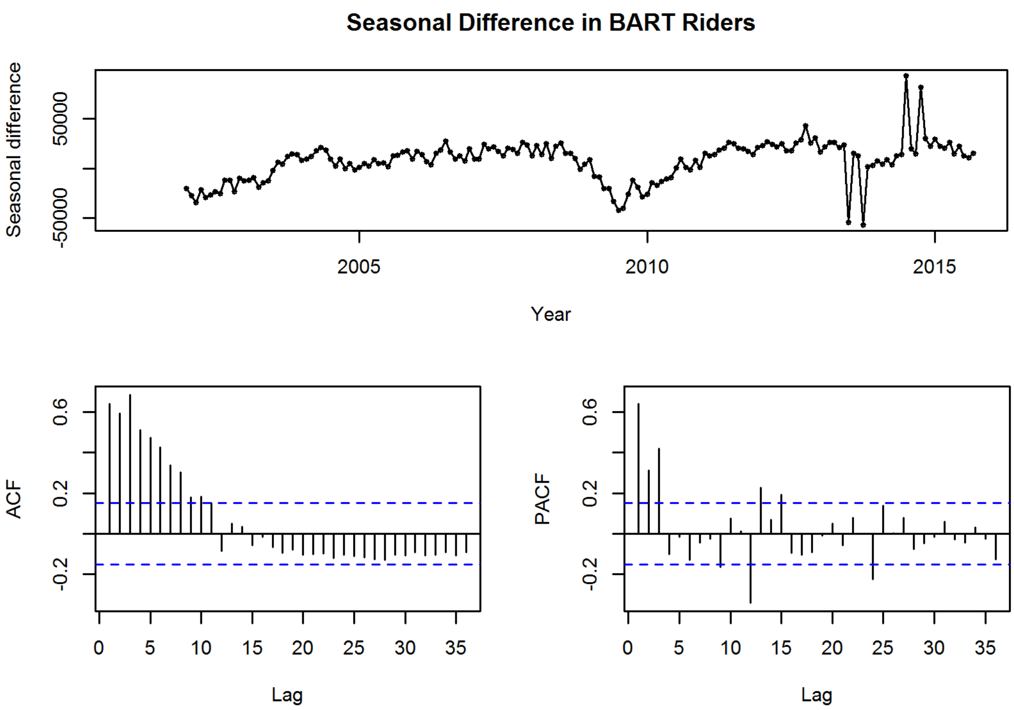


Figure 4.14: Autocorrelation of seasonal difference in BART riders

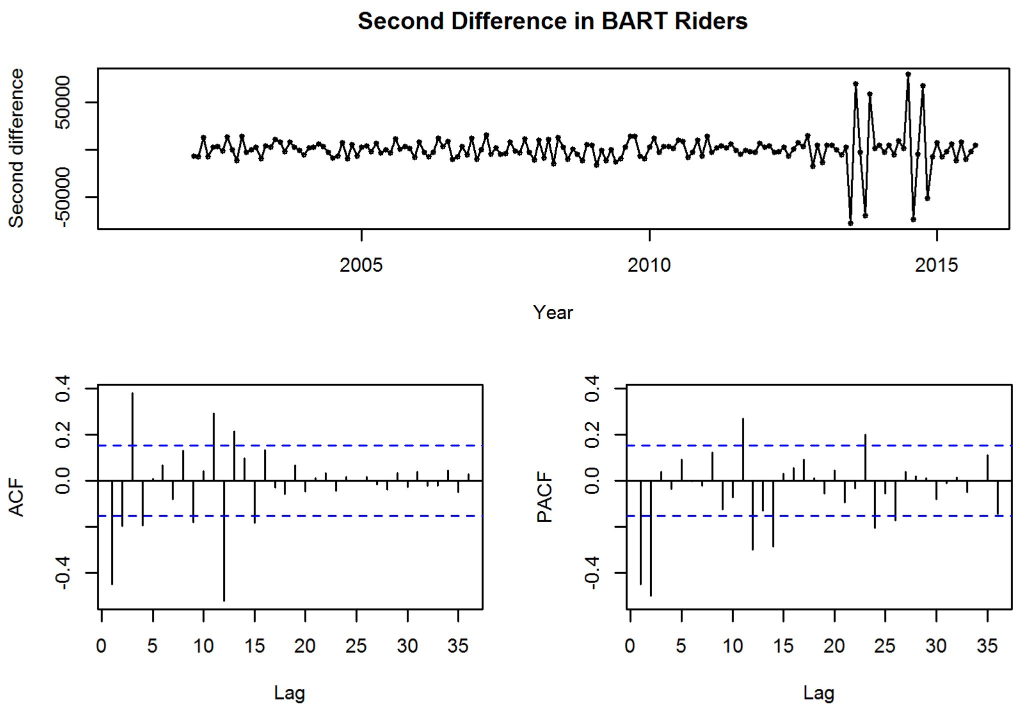


Figure 4.15: Autocorrelation of second difference in BART riders

4.4.2 ARIMA Models

This section shows the estimated ARIMA models of MUNI and BART ridership. These models aim simply to reproduce the patterns contained in the time series. While they do not offer explanation to why those patterns exist, they are included as a useful reference point in building towards more descriptive models.

4.4.2.1 MUNI ARIMA Models

Table 4.10 shows the estimation results for the preferred MUNI ARIMA model, which takes an $ARIMA(1,1,1)(0,1,0)_{12}$ form. In addition to both a seasonal difference and a monthly difference, the model includes a single autoregressive term (ϕ_1) and a single moving average term (θ_1). The lag column indicates that these apply to the values from the previous month. Both coefficients are statistically significant, and the goodness of fit measures are as shown.

Table 4.10: ARIMA model of MUNI boardings

Model Characteristics				
Dependent variable	MUNI boardings			
Type	$ARIMA(1,1,1)(0,1,0)_{12}$			
Date range	Jun 2009 to Nov 2013			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Autoregressive coefficient	1	0.4708	0.1751	2.69
Moving average coefficient	1	-0.9296	0.1051	-8.84
Model Statistics				
Log likelihood		-459.95		
AIC		925.90		
AICc		926.54		
RMSE		17782		
Percent RMSE		3.57%		
Box-Pierce test p-value		0.998		

Figure 4.16 shows ACF, PACF and time series plots for the residuals (e_t) from this model. Due to the second difference, model values cannot be calculated until after 13 periods. Therefore, the residuals are plotted as a horizontal line at value zero for the first 13 periods, allowing the bounds of the plot to remain consistent with other figures. The residuals appear stationary, and a Box-Pierce test, as reported in Table 4.10, produces a high p-value confirming their stationarity.

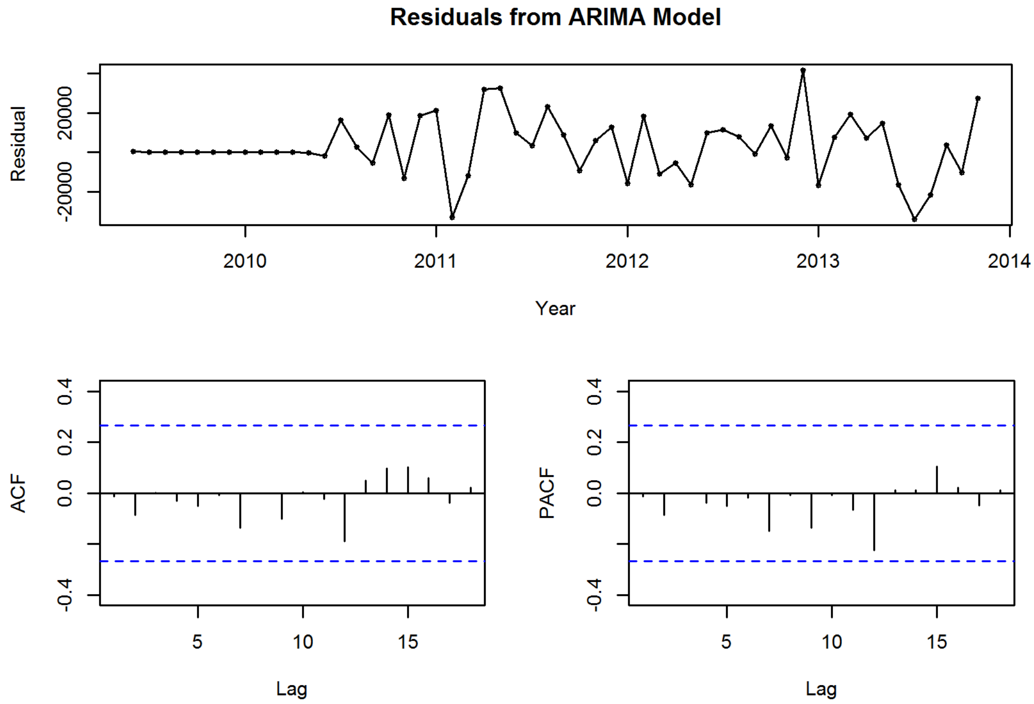


Figure 4.16: Residual autocorrelation from MUNI ARIMA model

Figure 4.17 shows a time series plot of the modelled and observed MUNI riders. The model is able to reproduce much of the observed pattern, although there are some noticeable deviations, particularly in spring 2011 and summer 2013.

In addition to time series plots, it is also valuable to examine the model's ability to replicate the actual change in the training data set. This is done here by comparing scatterplots for the modelled and observed annual difference and second difference, as shown in Figure 4.18. In each of these plots, the blue line shows the diagonal, and the points for a perfect model would lie on the diagonal. In this case, the changes are of the correct order of magnitude, but appear to be just a cloud. The RMSE for this model is 17,782, which is a 3.57% RMSE when calculated as a percent of the original series.

4.4.2.2 BART ARIMA Models

Table 4.11 shows the estimation results for the preferred BART ARIMA model, which takes an $ARIMA(2,1,0)(0,1,1)_{12}$ form. In addition to the second difference, the model includes two autoregressive terms (ϕ_1 and ϕ_2) and a seasonal moving av-

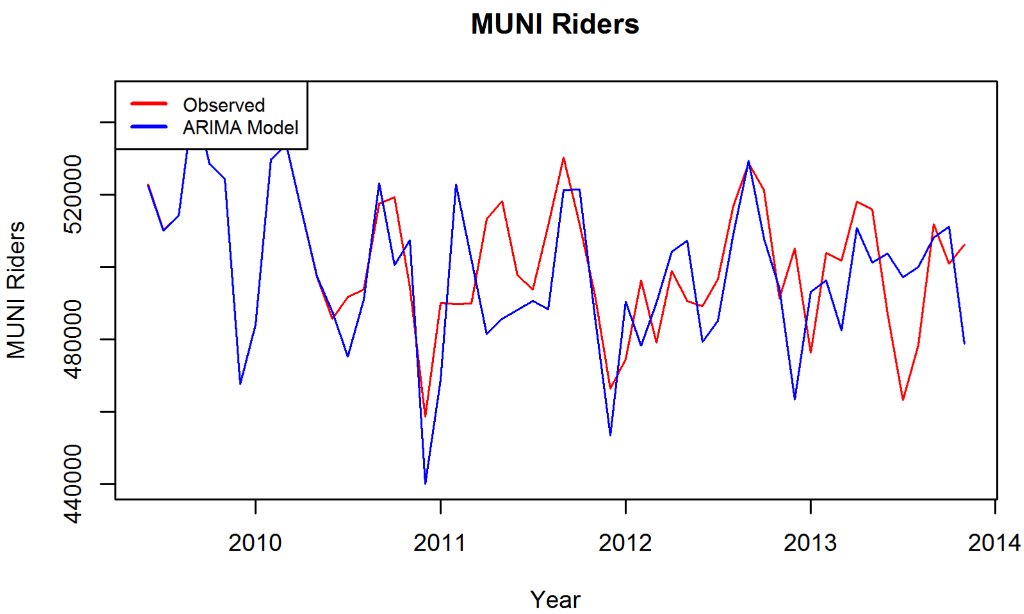


Figure 4.17: MUNI boardings, observed vs. ARIMA model

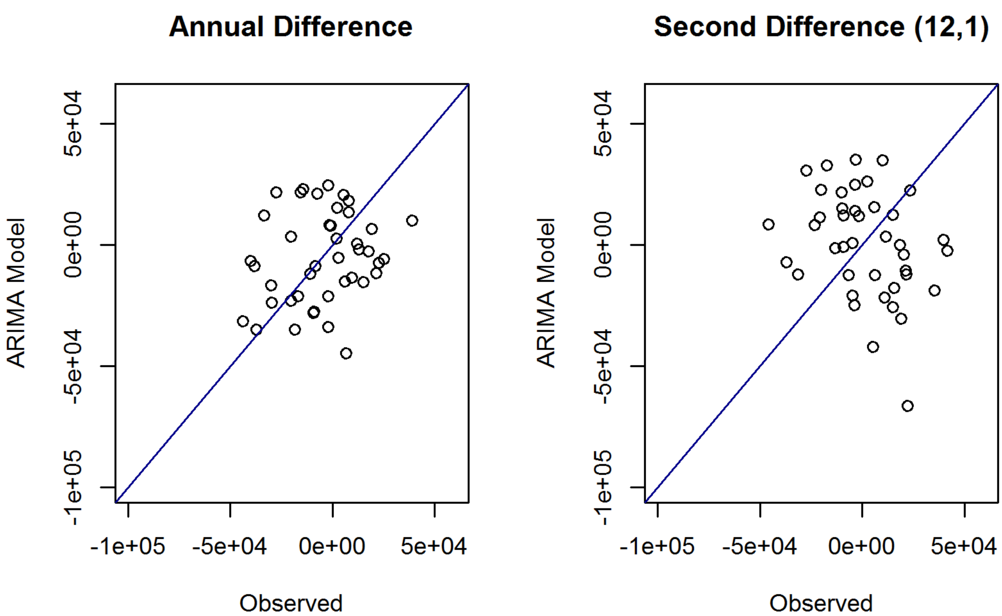


Figure 4.18: Change in MUNI boardings, observed vs. ARIMA model

erage term (Θ_1). The BART model is estimated over a much longer time series than the MUNI model, because there is a longer period over which data are available.

Table 4.11: ARIMA model of BART boardings

Model Characteristics				
Dependent variable	BART boardings			
Type	$ARIMA(2,1,0)(0,1,1)_{12}$			
Date range	Jan 2001 to Mar 2015			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Auto-regressive coefficient	1	-0.5977	0.0700	-8.54
Auto-regressive coefficient	2	-0.4494	0.0701	-6.41
Seasonal moving average coefficient	S1	-0.8631	0.0756	-11.42
Model Statistics				
Log likelihood		-1743.35		
AIC		3494.71		
AICc		3494.96		
RMSE		9508		
Percent RMSE		2.72%		
Box-Pierce test p-value		0.986		

Figure 4.19 shows ACF plots for the model's residuals, with non-zero values starting after 13 periods. These plots show that the residuals appear stationary, and a Box-Pierce test confirms that they are not autocorrelated.

Figure 4.20 shows a time series plot of the modelled and observed BART ridership. Figure 4.21 shows scatterplots of the modelled and observed seasonal and second differences. These appear to fit better than the MUNI model.

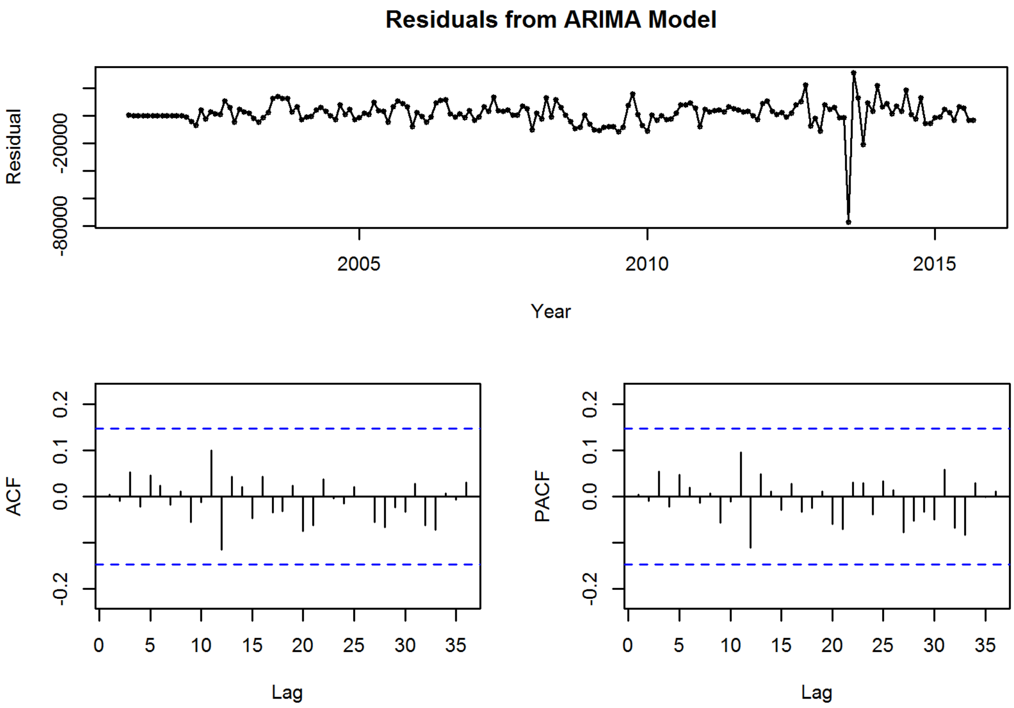


Figure 4.19: Residual autocorrelation from BART ARIMA model

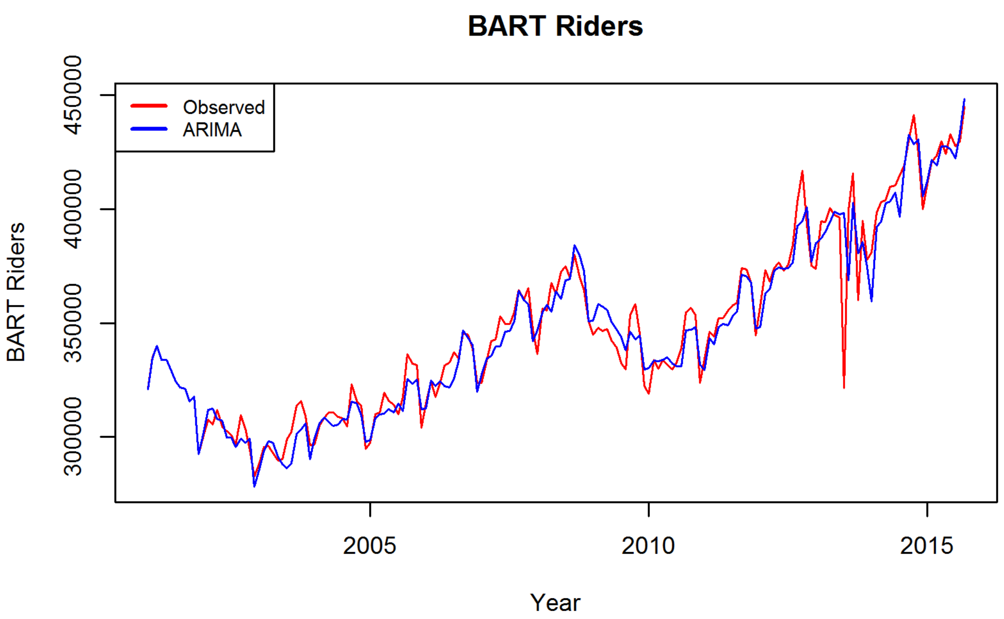


Figure 4.20: BART boardings, observed vs. ARIMA model

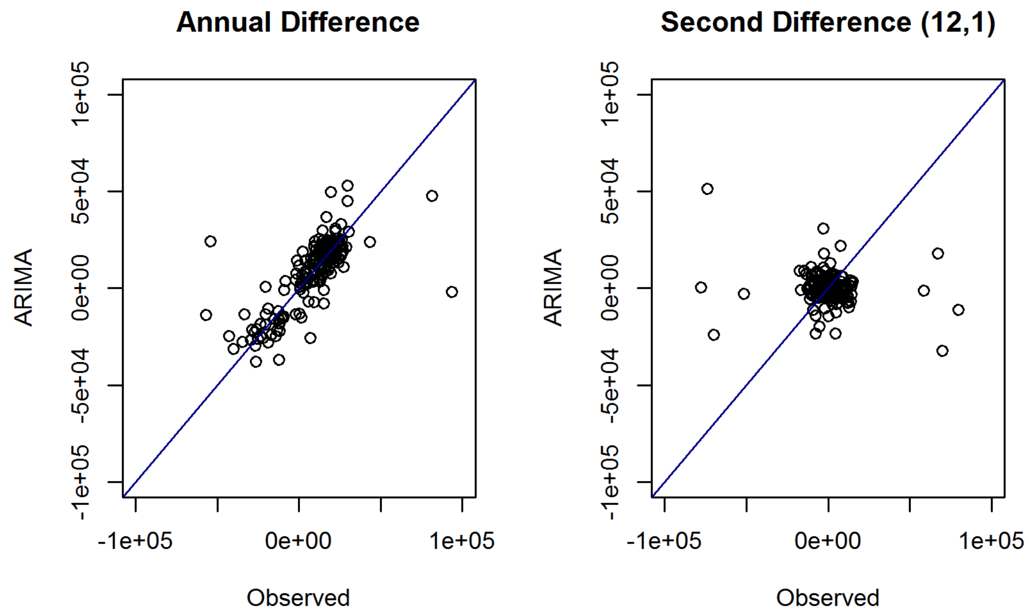


Figure 4.21: Change in BART boardings, observed vs. ARIMA model

4.4.3 Regression Models

The estimation results of the preferred regression models are presented here. Before estimating the models, both the dependent time series and all descriptive time series are transformed using a second difference, the first of which is an annual difference.

4.4.3.1 MUNI Regression Models

Table 4.12 shows the estimation results for the preferred regression model of average weekday MUNI boardings. The model is estimated from monthly data covering the period from June 2009 through November 2013. The model includes three variables that are statistically significant, and one that is marginally significant and included because it is a logical contribution to the model.

The weekday service miles is a key level-of-service measure, with a positive coefficient of 8,536 riders per 1,000 service miles. The service miles are measured from the GTFS data, and is specific to MUNI buses. There are several service increases and decreases during the estimation period.

In addition to the bus level-of-service, the model also represents competition with rail, by including the service miles on MUNI rail. The MUNI rail system is distinct from BART, and operates exclusively within San Francisco, so is more of a direct competitor. An increase of 1,000 rail service miles corresponds to a decrease of 4,352 bus riders. The bus is less sensitive to service changes on rail, than to service changes on bus, which is logical in the context of competition in the city transit system.

Next, the model includes the average bus runspeed. The model indicates that increasing the average runspeed by one mile per hour (mph) would increase ridership by 64,000. This would be a large improvement, as the average speed over the analysis period ranges between 10.5 and 11 mph.

Finally, the model includes the employment in San Francisco, with a coefficient of 2.2 additional MUNI boardings for each additional employee. For comparison, there are an average of about 500,000 MUNI boardings and 570,000 employees in San Francisco over this period, for an average rate of 0.88 boardings per employee. The estimated coefficient is higher than that, which may reflect that it is picking up some effect associated with ridership that is correlated with another factor. A number of specifications were tested in an effort to achieve a more pleasing model. This included the substitution of workers, population and households for employment, combinations of the three, and the separate inclusion of workers and non-workers. Employment was found to be the most reasonable specification.

Table 4.12: Regression model of MUNI boardings

Model Characteristics				
Dependent variable	MUNI boardings			
Type	Regression model of second differences			
Date range	Jun 2009 to Nov 2013			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Weekday service miles, 1000s		8536	3647	2.34
Weekday service miles on MUNI rail, 1000s		-4352	2725	-1.58
Average bus runspeed		63927	24592	2.60
Employment in San Francisco		2.201	0.953	2.31
Model Statistics				
R-squared		0.391		
RMSE		15,599		
Percent RMSE		3.13%		
Box-Pierce test p-value		0.816		

A number of additional variables were tested for inclusion in the model, but were rejected either because they gave insignificant parameter estimates, or because they resulted in coefficients that were illogical either in sign or in magnitude relative to related coefficients.

Some additional level-of-service specifications were tested, such as the substitution of average headway for service miles, and the inclusion of the number of stops or number of (bus) trip-stops as a measure of coverage. These were rejected.

The average bus total speed was tested as as a substitute for the average runspeed. Both are actual, not scheduled, speeds. They are derived from the AVL data, and processed as described in 3. Runspeed is the average speed between stops, measured from the time the door closes to the time the door opens. Total speed includes the dwell time at stops. Including the total speed produced a negative coefficient, meaning that higher speeds are correlated with fewer riders. This is the opposite of the effect that would be expected with respect to travel behaviour, and is thought to occur for operational reasons. That is, when ridership increases, the buses become more crowded, and it takes more time for the passengers to board and alight, slowing down the average total speed.

Measures were also tested for including transit crowding and reliability. The AVL/APC data allow both to be measured on a vehicle-by-vehicle bases, for those vehicles in the sample, with the values expanded to represent the system total. See 3 for a more detailed discussion of these calculations. In general, MUNI performs poorly on both measures, so with the ability to measure these as they relate to ridership, there should be an opportunity to estimate how much poor reliability and high crowding discourages transit patronage. Depending on the specification, a positive coefficient could be estimated for the average ontime performance, defined as the bus being no more than one minute early or five minutes late, but that estimate was insignificant. Some specifications resulted in a negative reliability coefficient, which may occur because more crowded buses tend to be operationally less reliable due to the time variation attributable to passenger boarding and alighting. Including a coefficient on crowding, defined as the share of bus trips where the number of passengers exceed 85% of the bus capacity, produces a model with a positive coefficient on crowding. This result is logical, but it is not helpful. It indicates that if there are more riders on the buses, then the buses will be more crowded, not that having more crowded buses would cause more people to ride the bus. Both these terms could be re-visited using panel data models, where there may be some variation across routes as well as through time, possibly resulting in more robust model estimates.

As noted above, combinations of population, households and workers were tested. Households were tested with segmentation both by income group and by auto ownership. The expectation was that lower income or lower auto households would be more likely to generate transit trips, but the estimation results did not show this to be the case. The median household income was tested as well, and this resulted in a positive and insignificant coefficient. Workers were tested with segmentation by worker earnings: with annual earnings up to $15,000, $15,000-$40,000, and over $40,000, but the additional segmentation did not improve the models. Median household income was tested, and the resulting coefficient was positive and insignificant. In many of these cases, the challenge may be that these income-related terms tend to be co-linear with employment, so the separate effects of the two may be hard to distinguish.

Several tests were done including employment segmented by different industries. In particular, it was thought that retail employment or health care employment might be important generators of transit trips, but this did not show up in the estimation results.

Cost terms were tested for inclusion as well, including both the cash fare and the average fare, considering monthly passes and other discounted fares. A reasonable coefficient could not be estimated on either, because the only notable fare change (beyond some relatively minor adjustments to monthly pass prices) to occur is only one month into the time series. The average car fuel cost per mile and average gasoline price per gallon were tested, but resulted in negative coefficients, indicating that higher fuel price would correspond to lower transit ridership. This relationship is the opposite of what we would expect.

Models were tested that included terms on the reported journey to work mode shares, derived from the American Community Survey data. The goal was to test whether the bike/taxi/other commuting, walking or working at home, which otherwise would not be well represented in this model structure, dampened transit demand. These tests resulted in negative parameter estimates, but the t-statistics were only about 1.5, depending somewhat on the remaining model specification, which did not provide sufficient confidence to include the terms in the model.

Distributed lags were tested on several variables, notably the level-of-service variables. The immediate response specification performed better than a distributed lag of three, six or 12 months.

Log specifications were tested on a number of variables, including the workers, employment and cost terms.

Figure 4.22 shows ACF, PACF and time series plots for the residuals from this model. These residuals appear stationary, and a Box-Pierce test, as reported in Table 4.12, produces a high p-value confirming their stationarity. As with the ARIMA models, the residuals from the regression model are zero for the first 13 periods due to the second difference.

Figure 4.23 shows the modelled and observed time series plots for this model. Compared to Figure 4.17, the regression model does a better job of replicating the observed values in summer 2013.

Figure 4.24 shows scatterplots comparing modelled and observed first and second differences. The clouds of points appears to be somewhat more consolidated than the MUNI ARIMA model, although still shows a high degree of scatter.

The RMSE for this model is 15,559, which is a 3.13% RMSE when calculated as a percent of the original series. This is an small improvement over the MUNI ARIMA model, which had an RMSE of 3.57%.

4.4.3.2 BART Regression Models

Table 4.13 shows the estimation results for the preferred regression model of average weekday BART ridership. The model is estimated from monthly data covering

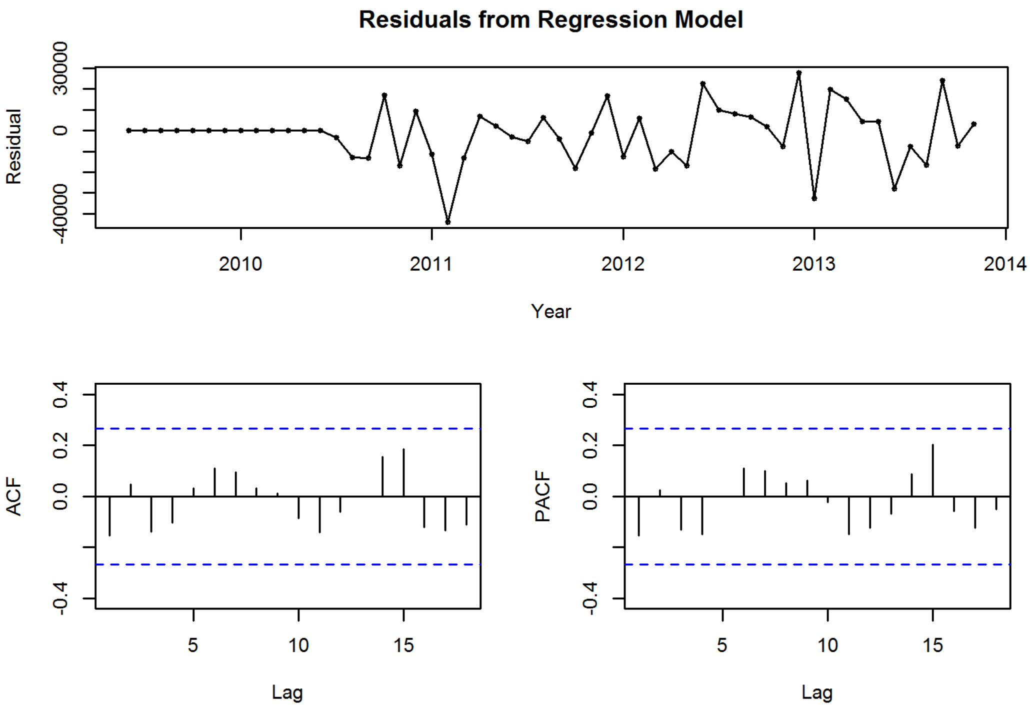


Figure 4.22: Residual autocorrelation from MUNI regression model

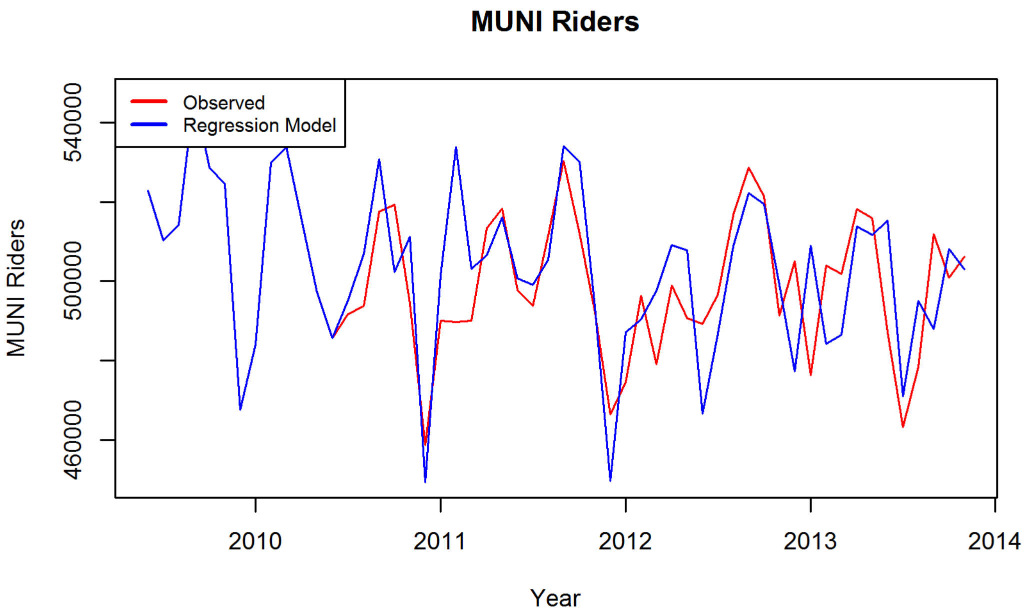


Figure 4.23: MUNI boardings, observed vs. regression model

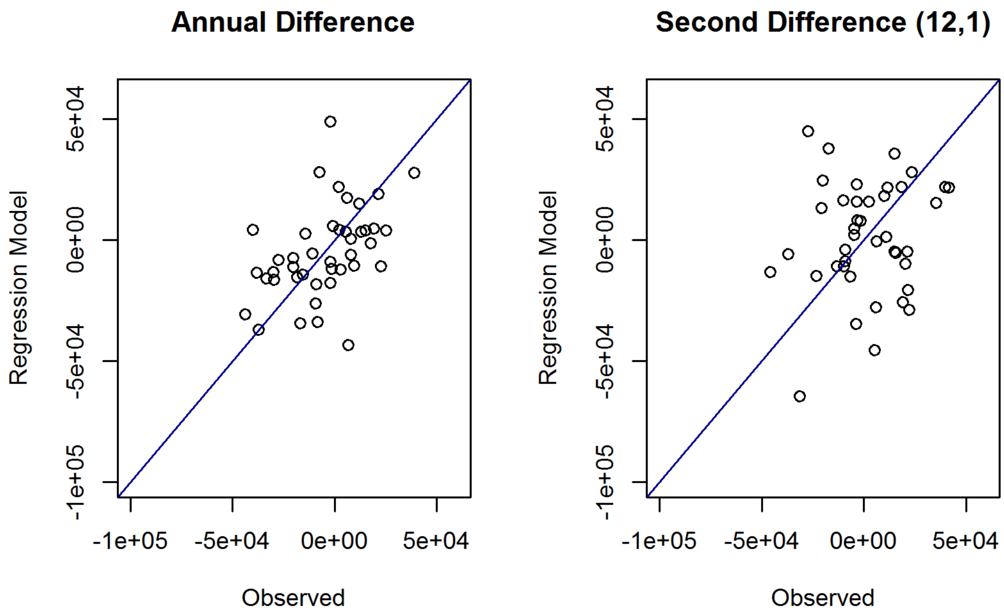


Figure 4.24: Change in MUNI boardings, observed vs. regression model

the period from January 2001 through March 2015. The model includes four variables, all of which are statistically significant.

The only level-of-service variable found to be significant is the number of stations. The number of stations increased from 39 to 43 in 2003 when the extension to San Francisco Airport opened. In 2011, an infill station opened at West Dublin/Pleasanton, and in 2014 a connection opened to Oakland Airport. The results indicate that each additional station corresponds to an increase of 7,613 riders. Several lags were tested for this variable, and a 12-month distributed lag was found to work best. This implies that it takes travellers a year to adjust their travel patterns to the new infrastructure. For each month in the first year of a new station, the model would predict an increase of 634 riders (7,613 / 12). Service miles were also tested as a level-of-service measure, but the parameter estimates were insignificant in the regression models.

Next, the model includes employment, with each additional employee in the 4-county area generating 0.18 BART riders. This is much lower than the number of MUNI boardings generated per employee because the area is larger, and more travellers commute by car or other modes.

The average cash fare is also included. The model indicates that a one dollar increase in the fare would cause a reduction of about 23,000 riders.

BART employee strikes occurred in July and October 2013, while a new labour contract was under negotiation [203, 204]. The first lasted four days, and the second lasted three. These two months are noticeable outliers, with lower ridership than the months before and after. Including a variable for the number of days in the month with a BART strike resulted in a much improved model in terms of goodness of fit, and a highly significant parameter estimate. This specification performed better than a binary flag of whether or not a strike occurred, because it reflects the larger magnitude of the July strike. The estimate indicates that the average weekday ridership for the month is 20,000 less for each additional day of strikes.

Table 4.13: Regression model of BART boardings

Model Characteristics				
Dependent variable	BART boardings			
Type	Regression model of second differences			
Date range	Jan 2001 to Mar 2015			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Number of Stations	D(0,12)	7613	3418	2.23
Employment in 4-county area		0.1827	0.0522	3.50
Cash fare (2010 $)		-23490	10250	-2.29
Days with a BART strike		-19690	684.5	-28.77
Model Statistics				
R-squared		0.845		
RMSE		6733		
Percent RMSE		1.94%		
Box-Pierce test p-value		2.368e-09		

In addition to these terms included in the model, a number of additional variables were considered for inclusion in the models. As with MUNI, combinations of population, households and workers were tested, including tests of household by income group, workers by earnings, and employment by industry. The employment specifically in San Francisco County and the share of employment in San Francisco County were tested, but not found to be distinct and significant relative to total employment. Average fuel cost per mile, average gasoline cost per gallon and median household income were tested, but did not improve the model significantly.

The level-of-service terms available for the BART estimation were more limited than for the MUNI estimation, because the data are less detailed. The demand data are available as entry-exit matrices showing where travellers enter and exit the system's faregates, but detailed AVL/APC data is not available at the level of individual vehicles, as it is for MUNI. The GTFS data do provide a detailed enumeration of the scheduled service, but is only available as far back as 2009. Prior to 2009, only an aggregate reporting of the service miles is available.

Detailed crowding and reliability are not available for BART. Anecdotally, reliability is much less of a problem on BART than on MUNI, although crowding is an issue, with the transbay tube currently at capacity during peak periods [205].

The cost of tolls and carpool policies on the Bay Bridge were tested as predictive variables, but not found to be significant.

Models were tested that included terms on the reported journey to work mode shares, but these did not relate to BART ridership in the same way they did to MUNI ridership. Distributed lags were tested on the cost and employment terms. Log specifications were tested on a number of variables, including the workers, employment and cost terms.

Figure 4.25 shows ACF, PACF and time series plots for the residuals from this model. Non-zero values start at period 13 due to the second difference. These show significant autocorrelation for several lags, notably at months one and 12. A Box-Pierce test, as reported in Table 4.13, produces a very small p-value, confirming these autocorrelations. This result is problematic, because it indicates that the regression results could be spurious.

Figure 4.26 shows a time series plot of the modelled and observed BART ridership. Figure 4.27 shows scatterplots of the modelled and observed seasonal and second differences.

The RMSE for this model is 6,733, which is a 1.94% RMSE when calculated as a percent of the original series. This is an improvement over the BART ARIMA model, which had an RMSE of 2.72%. In spite of the improved fit, the model is rejected due to the risk of spurious regression, and RegARIMA models are explored

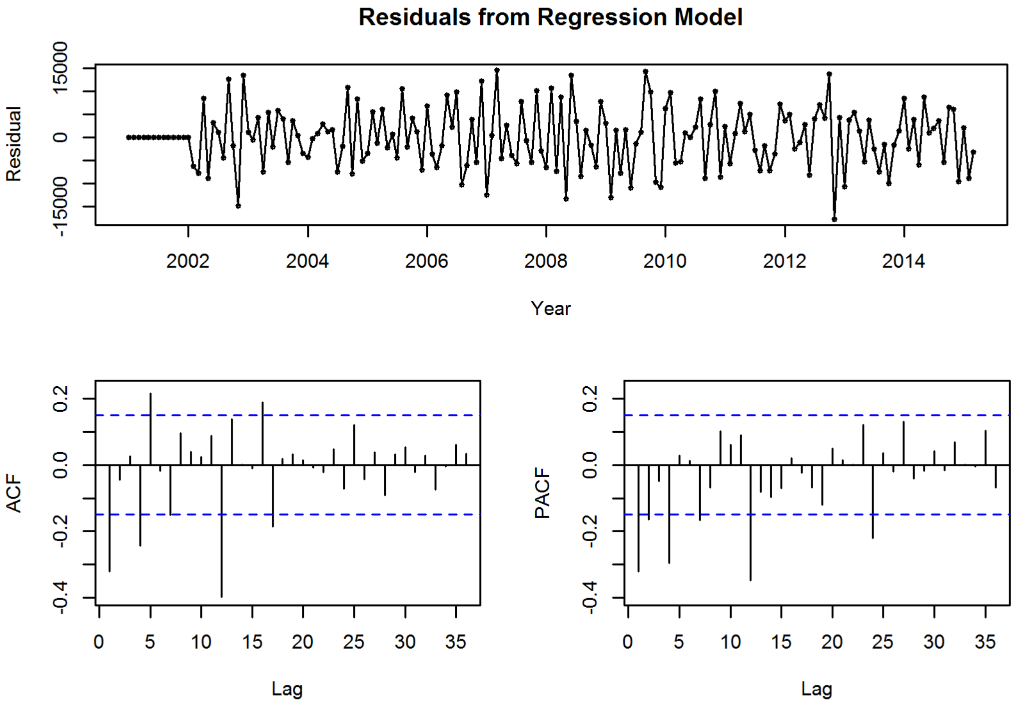


Figure 4.25: Residual autocorrelation from BART regression model

as an alternative.

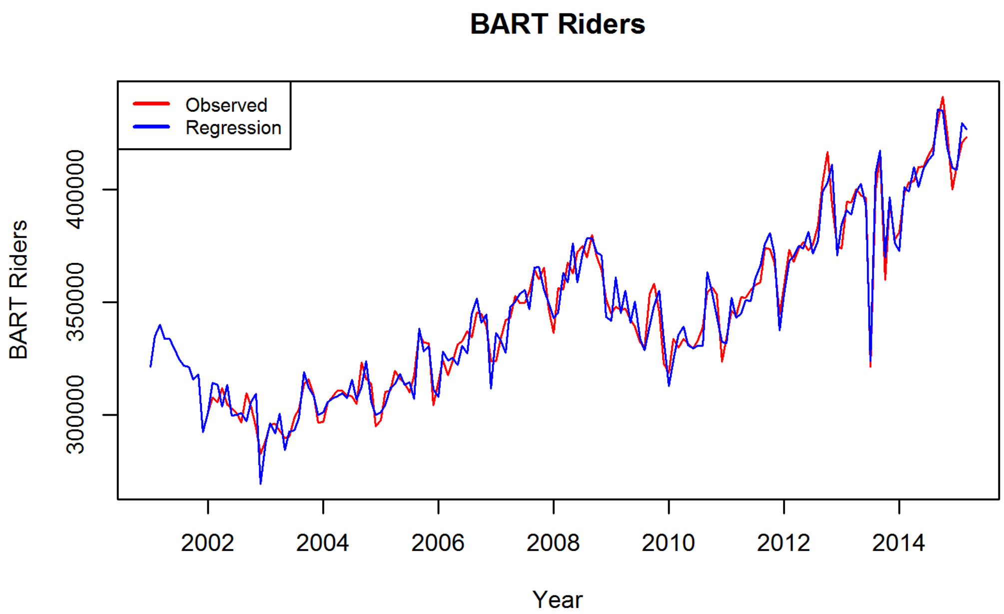


Figure 4.26: BART boardings, observed vs. regression model

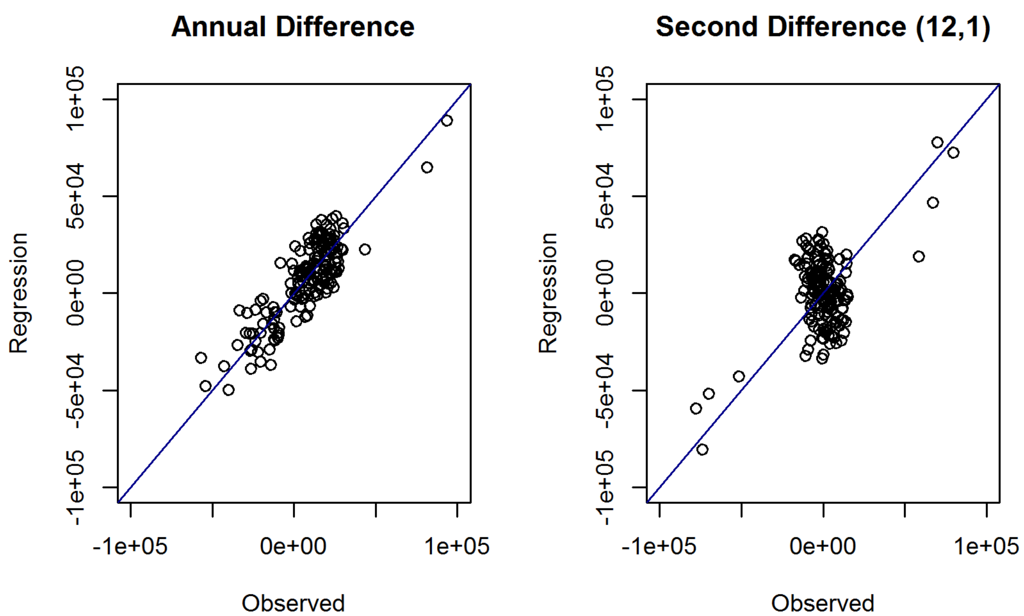


Figure 4.27: Change in BART boardings, observed vs. regression model

4.4.4 Regression Models with ARIMA Errors

The estimation results for the RegARIMA models are presented here.

4.4.4.1 MUNI RegARIMA Models

Table 4.14 shows the estimation results for the preferred RegARIMA model of average weekday MUNI boardings. The model is estimated from monthly data covering the period from June 2009 through November 2013.

The ARIMA portion of the model takes the form $ARIMA(0,1,0)(0,1,0)_{12}$. The autoregressive and moving average terms drop out of the model, reducing it to a regression model on data that has been transformed with a second difference. In fact, it is the same model as the preferred regression model described above. The method of estimation being the only difference.

Table 4.14: RegARIMA model of MUNI boardings

Model Characteristics				
Dependent variable	MUNI boardings			
Type	$ARIMA(0,1,0)(0,1,0)_{12}$			
Date range	Jun 2009 to Nov 2013			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Weekday service miles, 1000s		8536	3647	2.34
Weekday service miles on MUNI rail, 1000s		-4352	2725	-1.58
Average bus runspeed		63927	24592	2.60
Employment in San Francisco		2.201	0.953	2.31
Model Statistics				
Log likelihood		-454.03		
AIC		918.06		
AICc		919.78		
RMSE		15,599		
Percent RMSE		3.13%		
Box-Pierce test p-value		0.691		

A similar set of variables were tested and rejected in the RegARIMA models as in the regression models. The goal was to determine whether a more flexible model structure would allow the model to pick up more or different correlations. While the results were different for a number of specifications, none was preferred over this specification.

Several additional variables were tested in the RegARIMA models that were not previously tested in the regression models. These variables focus on the possible effect of car ownership, car congestion and car mode shares.

The possible effect of car ownership focused specifically on households that own zero vehicles, because those households are the most constrained in their travel behaviour and therefore more likely to use transit. Car ownership data are available annually at a county level from the ACS. Two variables were tested: the absolute number of zero-car households, and the share of households that are zero-car. The coefficient on the total number of zero-car households was positive, but insignificant. The coefficient on the share of households that are zero-car was negative and insignificant.

Two aggregate measures of congestion were considered in the region: the Texas A&M Transportation Institute (TTI) travel time index and the TomTom traffic index. The TTI travel time index is the ratio of travel time in the peak period to travel time in free-flow conditions. It increases from 1.38 in 2009 to 1.40 in 2013. The estimated coefficient is negative, but insignificant. The TomTom traffic index is a measure of the extra travel time associated with congestion. It increases from 26% in 2009 to 32% in 2013. The estimated coefficient is positive, but insignificant.

Car mode shares were measured using the ACS journey-to-work commute mode share data. The analysis considers both the drive-alone mode share and the total car mode share (drive-alone plus carpool). Both generally decrease over the analysis period, but not uniformly, which may indicate some amount of noise in the data. The drive-alone mode shares in 2009 through 2013 are: 38.9%, 36.0%, 37.7%, 36.3% and 36.4%. The total car commute mode shares from 2009 through 2013 are: 46.4%, 43.9%, 45.0%, 44.0% and 43.2%. The estimated coefficients on both are large and negative (a 1% decrease in the car mode share would be associated with a increase of 18,000 MUNI bus riders), but insignificant.

These tests of car ownership, car congestion and car mode shares do not lead us to select a different preferred model.

Figure 4.28 shows ACF, PACF and time series plots for the residuals from this model. These residuals appear stationary, and a Box-Pierce test, as reported in Table 4.14, produces a high enough p-value to confirm the lack of significant autocorrelation. Figure 4.29 shows the modelled and observed time series plots for this model. Figure 4.30 shows scatterplots comparing modelled and observed first and second differences.

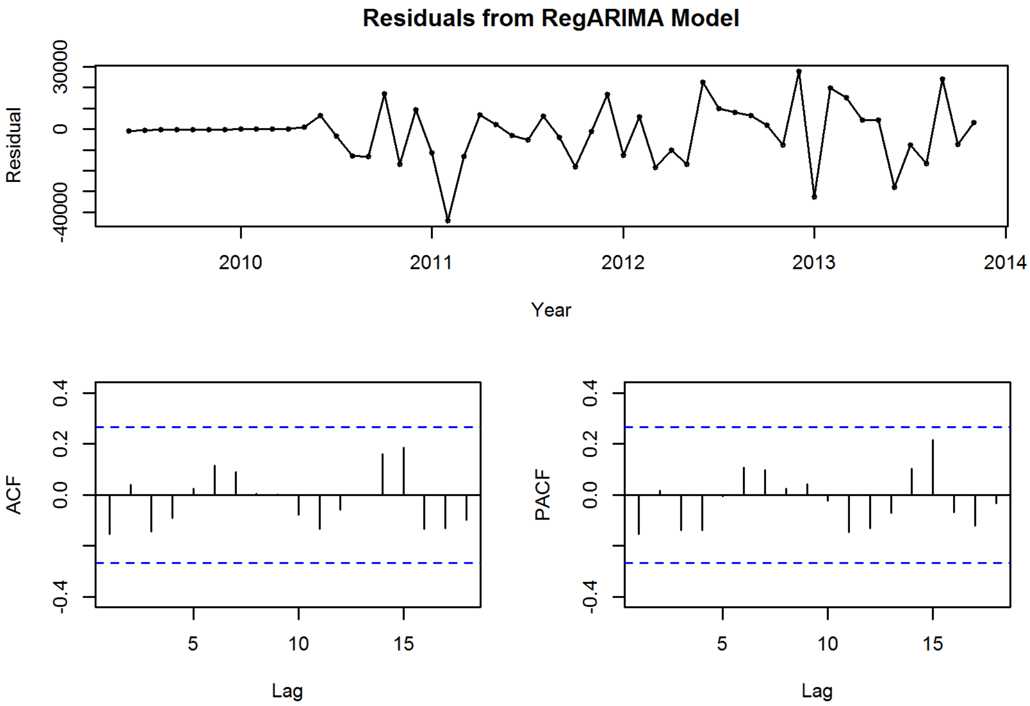


Figure 4.28: Residual autocorrelation from MUNI RegARIMA model

4.4.4.2 MUNI RegARIMA with Constrained Employment Term

Recognising that the estimated MUNI models are more sensitive to changes in employment than might be expected, an alternative set of models is estimated in which the employment coefficient is constrained. The constrained coefficient is set to to the average rate of MUNI boardings per employee over the analysis period, or 0.876, and the remaining model coefficients are re-estimated.

Table 4.15 shows the results of this estimation. The same exogenous terms are included in the model, but the selected structure becomes $ARIMA(0,1,1)(0,1,0)_{12}$, with the addition of a moving average coefficient. The signs of all the coefficients remain the same, but the magnitude change somewhat, and the t-statistic is lower for rail service in particular.

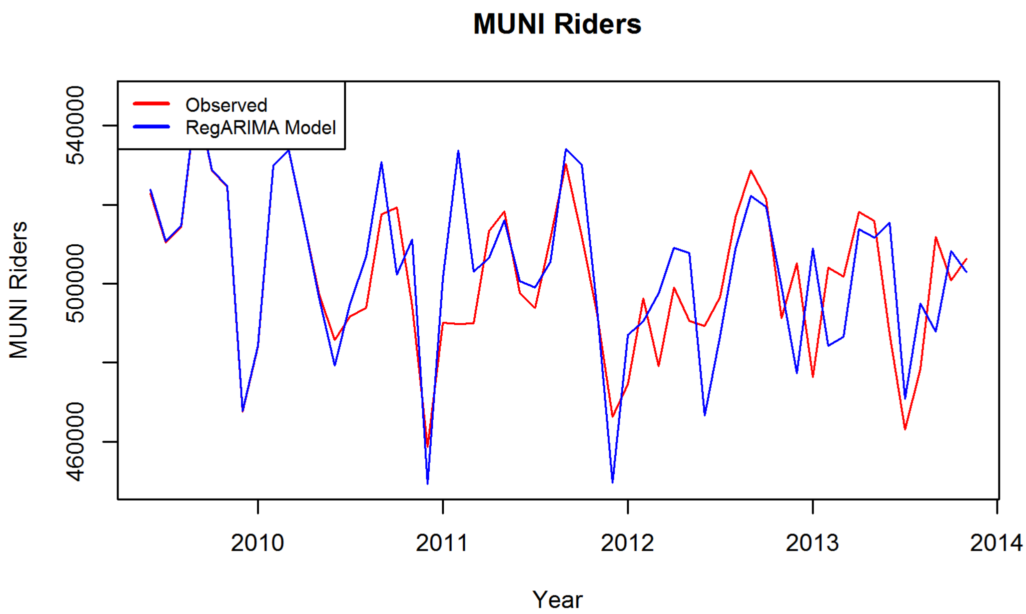


Figure 4.29: MUNI boardings, observed vs. RegARIMA model

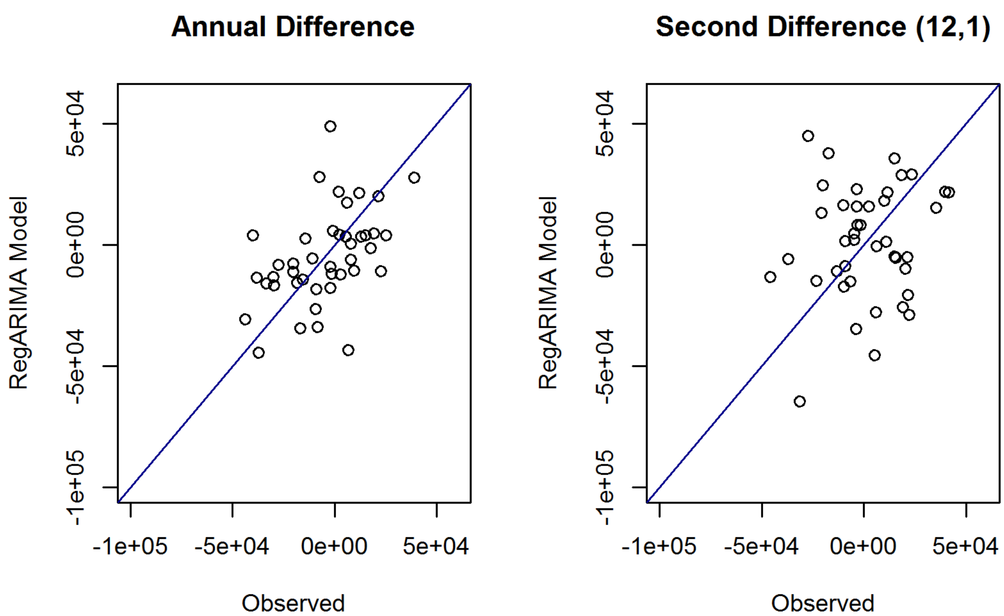


Figure 4.30: Change in MUNI boardings, observed vs. RegARIMA model

Table 4.15: RegARIMA model of MUNI boardings with constrained employment term

Model Characteristics				
Dependent variable	MUNI boardings			
Type	$ARIMA(0,1,1)(0,1,0)_{12}$			
Date range	Jun 2009 to Nov 2013			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Moving average coefficient	1	-0.3092	0.1852	-1.67
Weekday service miles, 1000s		7971	3105	2.57
Weekday service miles on MUNI rail, 1000s		-2777	2488	-1.12
Average bus runspeed		49853	25692	1.94
Employment in San Francisco		0.876	fixed	fixed
Model Statistics				
Log likelihood		-453.56		
AIC		917.12		
AICc		918.83		
RMSE		15,401		
Percent RMSE		3.09%		
Box-Pierce test p-value		0.873		

As with the other models, a variety of model specifications were tested, but a better one could not be found.

The model residuals are confirmed to be stationary, based on an inspection of the ACF and PACF plots in Figure 4.31, and the Box-Pierce test shown in Table 4.15. Figure 4.32 shows the modelled and observed time series plots, and Figure 4.33 shows scatterplots.

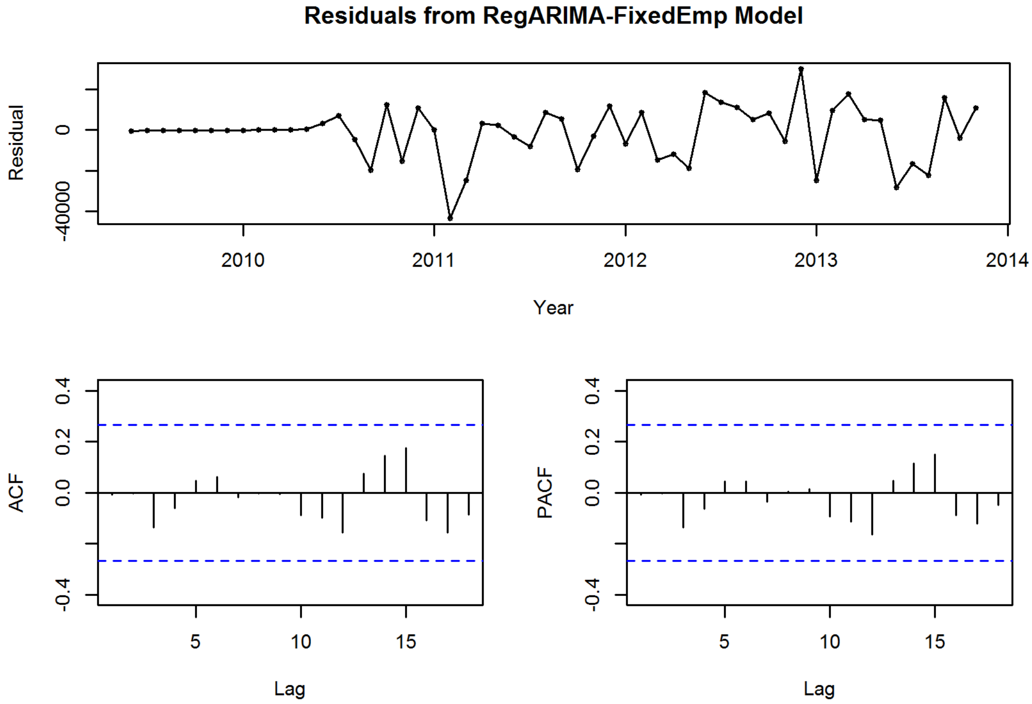


Figure 4.31: Residual autocorrelation from MUNI RegARIMA model with constrained employment term

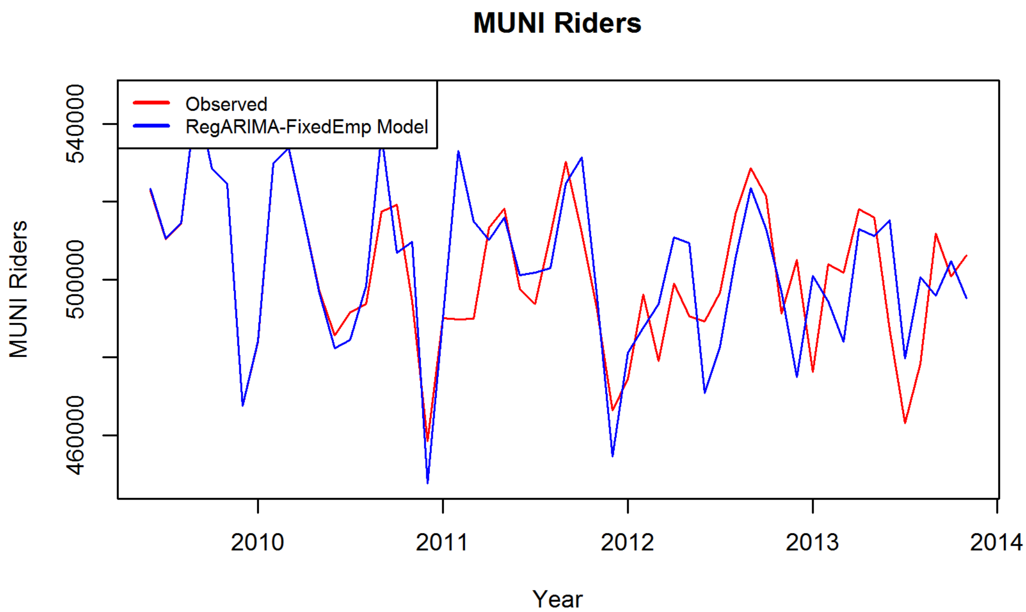


Figure 4.32: MUNI boardings, observed vs. RegARIMA model with constrained employment term

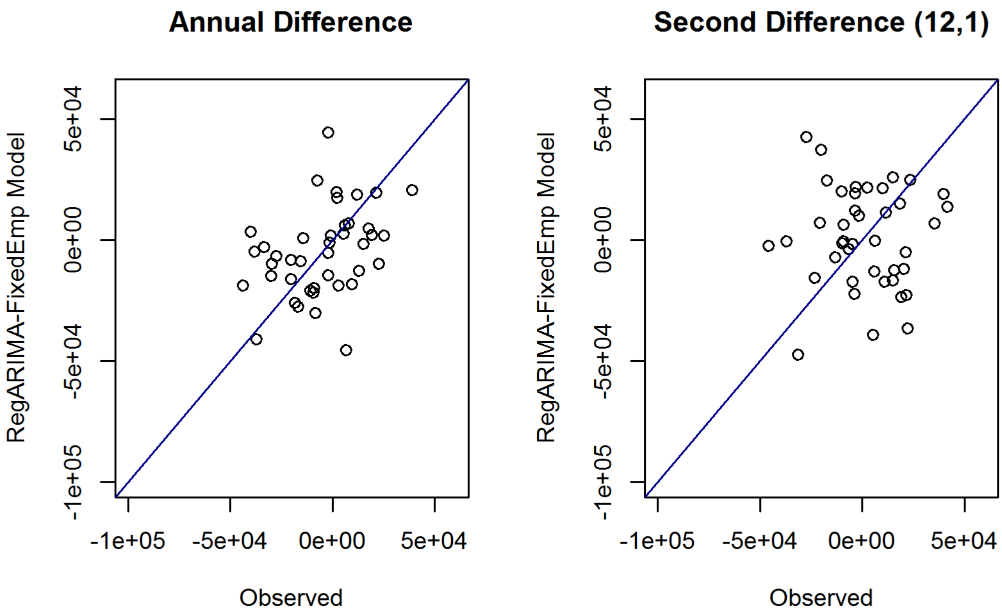


Figure 4.33: Change in MUNI boardings, observed vs. RegARIMA model with constrained employment term

4.4.4.3 BART RegARIMA Models

Table 4.16 shows the estimation results for the preferred RegARIMA model of average weekday BART boardings. The model is estimated from monthly data covering the period from January 2001 through March 2015.

The ARIMA portion of the model takes the form $ARIMA(0,1,2)(0,1,1)_{12}$. There is a second difference, two moving average terms, and a seasonal moving average term. This structure offers a mechanism to capture the residual autocorrelation found in the BART regression model.

The model includes all of the variables already included in the regression model, plus three additional terms: service miles, the percent of employment in San Francisco and the average car fuel cost.

The service miles coefficient, which was found in the regression model to be insignificant, shows up significant and with a logical sign. The model implies that each increase of 1,000 service miles results in 2,712 additional riders. Also, each new station results in 5,472 additional riders. Both terms are specified with a 12-month distributed lag, meaning that when there is a change, 1/12th of the additional ridership is allocated each month for a year. Several lag specifications were tested, and this worked best. It is interesting to note that testing distributed lags on the MUNI service miles was not effective. This could be because the MUNI models are estimated from a shorter time series, making it more difficult to capture lagged effects.

The model results indicate that each additional employee in the 4-county area generates 0.2 additional BART riders. This is similar in magnitude to the regression models, but much more strongly significant.

A coefficient was estimated on the percent of 4-county employment in San Francisco. This coefficient is positive, and indicates that an increase of 1% of the employment in San Francisco corresponds to an increase of about 5,500 BART riders. This term represents the employment concentration, as distinct from the total employment. Commute trips to downtown San Francisco are a core market for BART, so it makes sense that BART ridership is closely related to San Francisco County employment specifically.

The cash fare shows up as negative and significant, with each dollar increase in fare corresponding to a decrease of about 21,000 riders.

In addition to the fare, a significant coefficient could be estimated for the average car fuel cost in this model. It is positive, which means that higher costs to drivers result in a mode shift to BART. The fuel cost is expressed in dollars per mile, and is based on the price per gallon of gasoline and the average fuel economy of the fleet. Average fuel economy increases over this period, dampening the effect of increases in gasoline price.

The number of days with a BART strike is negative, significant, and of a similar magnitude to the regression models. The effect of this term is similar to what would be achieved by estimating the models only from the average ridership on non-strike weekdays during those months.

As with the BART regression models, a number of terms were tested for inclusion in the model, but rejected. These include various combinations of population, households, workers and employment segmented by income, earnings or industry.

Table 4.16: RegARIMA model of BART boardings

Model Characteristics				
Dependent variable	BART boardings			
Type	$ARIMA(0,1,2)(0,1,1)_{12}$			
Date range	Jan 2001 to Mar 2015			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Moving average coefficient	1	-0.5701	0.1122	-5.08
Moving average coefficient	2	-0.2827	0.1032	-2.74
Seasonal moving average coefficient	S1	-0.6603	0.0782	-8.44
Weekday service miles, 1000s	D(0,12)	2712	1310	2.07
Number of Stations	D(0,12)	5472	1057	5.18
Employment in 4-county area		0.2027	0.0185	10.96
Percent of 4-county employment in SF		8099	3860	2.10
Cash fare (2010 $)		-20795	8332	-2.50
Average car fuel cost (2010 $/mile)		86312	31504	2.74
Days with a BART strike		-19010	906.5	-20.97
Model Statistics				
Log likelihood		-1571.65		
AIC		3165.30		
AICc		3167.04		
RMSE		4923		
Percent RMSE		1.42%		
Box-Pierce test p-value		0.2068		

Tests of the median household income showed it to be positively correlated with BART ridership, although not significant.

Several different specifications were tested for how to include both total employment and employment in San Francisco. A challenge is that they tend to be co-linear both because employment in San Francisco is a large share of employment in the 4-county area, and because they are both related to the state of the economy. The total employment and percent in San Francisco worked well because it provided both a measure of the quantity and the concentration.

As with the regression models, the level-of-service variables available for inclusion in the BART RegARIMA models are somewhat limited, although the combination of service miles and stations appears to work reasonably well.

In contrast to the MUNI RegARIMA models, BART ridership does not appear to be related to the journey to work mode shares for walk, bike, taxi and other. This may be because BART trips tend to be longer, and often across the Bay, so there is less opportunity for substitution.

Several alternatives were tested that included either distributed lags or log transformations of variables.

As with the MUNI models, several additional variables were tested in the BART RegARIMA models to consider the possible effects of car ownership, car congestion and car mode shares.

The estimated coefficients on both the total number and the share of zero-vehicle households were negative and insignificant. The drive-alone mode share and the car mode share coefficients were both positive and insignificant. The TTI travel time index produced a coefficient that was positive (indicating more congestion is associated with higher BART ridership), but insignificant. The TomTom traffic index only goes as far back as 2008, so an estimated model with that as a variable must be based on a shorter time series than the full BART data set that goes back to 2001. This was done as a test, and it produced a coefficient on the TomTom traffic index that was marginally significant, with a t-statistic of 1.74. However, the coefficient was negative, indicating that higher levels of congestion are associated with lower BART ridership: the opposite of what we would expect. In addition, other coefficients in the model change substantially and become insignificant when estimated using this shorter time series. Given these results, the preferred models remain unchanged.

Figure 4.34 shows ACF, PACF and time series plots for the residuals from the BART RegARIMA model. These plots show a possible autocorrelation for a lag of 17 months, but the Box-Pierce test, as reported in Table 4.16, has a p-value of 0.2, indicating that the autocorrelations as a whole are not significant.

Figure 4.35 shows a time series plot of the modelled and observed BART ridership. The model appears to track the observed ridership nicely. Figure 4.36 shows scatterplots of the modelled and observed seasonal and second differences. The RMSE for this model is 4,923, which is a 1.42% RMSE when calculated as a percent of the original series. This is an improvement over the regression model, which had a 1.94% RMSE.

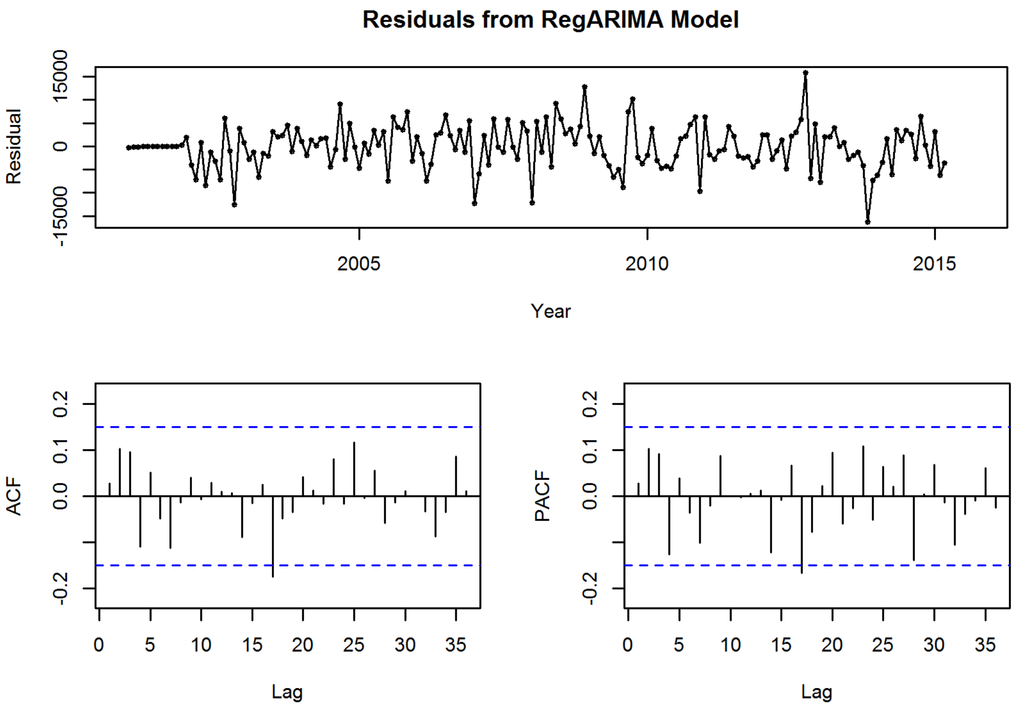


Figure 4.34: Residual autocorrelation from BART RegARIMA model

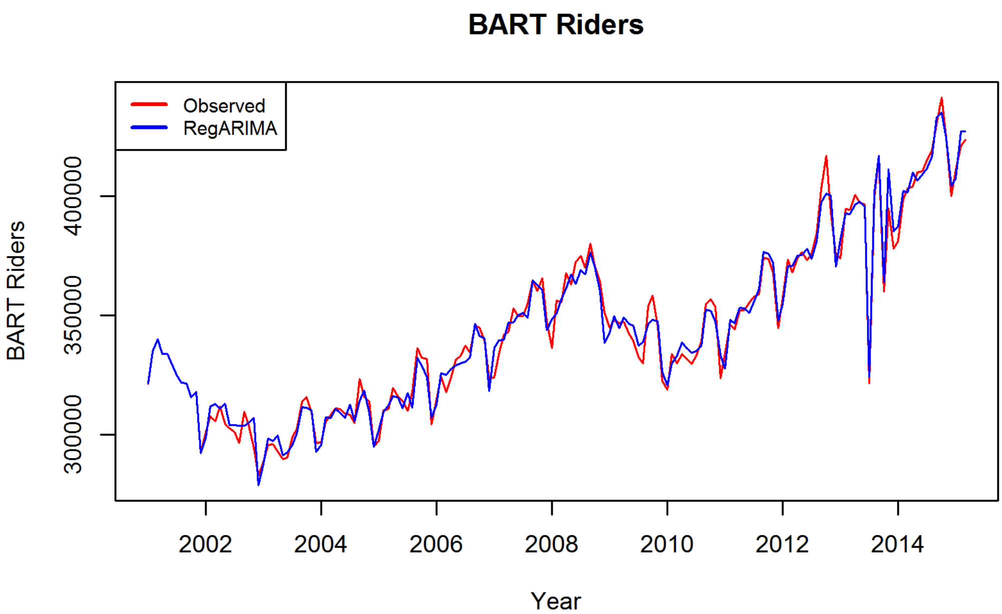


Figure 4.35: BART boardings, observed vs. RegARIMA model

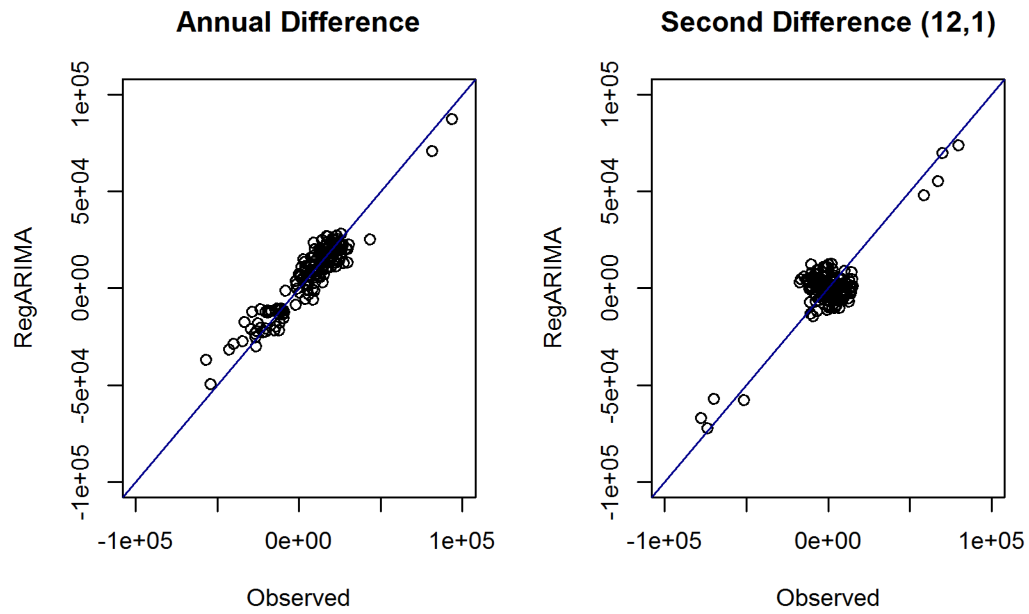


Figure 4.36: Change in BART boardings, observed vs. RegARIMA model

4.4.5 Comparison of Results

Having come to a preferred model of each type, the three models are compared here.

4.4.5.1 MUNI Comparison

Table 4.17 compares the estimation results for the three models of MUNI ridership. There is little to compare between the ARIMA models and the others, and the other two models are the same. The regression and RegARIMA models have a lower RMSE than the ARIMA model.

Table 4.18 compares the RegARIMA model with an estimated employment coefficient to the RegARIMA model with a constrained employment coefficient. In comparing these models the most prominent difference is the employment term itself, as well as the introduction of a moving average coefficient. The service miles coefficient is of the same sign, and similar in magnitude. The rail service coefficient is still negative, but about 40% lower in magnitude. The bus runspeed coefficient is still positive, but about 20% lower in magnitude.

The goodness of fit measures are slightly better in the constrained model. This

Table 4.17: Three models of MUNI boardings

Model Characteristics							
Dependent variable	MUNI boardings						
Date range	Jun 2009 to Nov 2013						
Predictive Variables							
		ARIMA		Regression		RegARIMA	
Description	Lag	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
Autoregressive coefficient	1	0.4708	2.69				
Moving average coefficient	1	-0.9296	-8.84				
Weekday service miles, 1000s				8536	2.34	8536	2.34
Weekday service miles on MUNI rail, 1000s				-4352	-1.58	-4352	-1.58
Average bus runspeed				63927	2.60	63927	2.60
Employment in San Francisco				2.201	2.31	2.201	2.31
Model Statistics							
RMSE		17,782		15,599		15,599	
Percent RMSE		3.57%		3.13%		3.13%	
Box-Pierce test p-value		0.998		0.691		0.691	

is presumably because there is one extra parameter included (the moving average coefficient), but the number of parameters actually estimated is the same.

Table 4.18: Comparison of MUNI RegARIMA Models with and without employment constraint

Model Characteristics					
Dependent variable	MUNI boardings				
Date range	Jun 2009 to Nov 2013				
Predictive Variables					
		Unconstrained		Constrained	
Type		$ARIMA(0,1,0)(0,1,0)_{12}$		$ARIMA(0,1,1)(0,1,0)_{12}$	
Description	Lag	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1			-0.3092	-1.67
Weekday service miles, 1000s		8536	2.34	7971	2.57
Weekday service miles on MUNI rail, 1000s		-4352	-1.58	-2777	-1.12
Average bus runspeed		63927	2.60	49853	1.94
Employment in San Francisco		2.201	2.31	0.876	fixed
Model Statistics					
Log likelihood		-454.03		-453.56	
AIC		918.06		917.12	
AICc		919.78		918.83	
RMSE		15,599		15,401	
Percent RMSE		3.13%		3.09%	
Box-Pierce test p-value		0.691		0.873	

A reasonable case can be made for selecting either of these models. Our tendency is to prefer the model with a constrained employment term, because it results in a model whose elasticity with respect to changes in employment is more in line with other published results, as examined in Chapter 5.

4.4.5.2 BART Comparison

Table 4.19 compares the estimation results for the three models of BART ridership.

The ARIMA model includes two autoregressive terms and one seasonal moving average term, whereas the RegARIMA model includes two moving average terms and one seasonal moving average term.

The RegARIMA model includes a broader array of descriptive variables than the regression model, including the weekday service miles, the percent of 4-county employment in San Francisco, and the average car fuel cost. These terms were tested in the regression model, but were found to be insignificant or otherwise unsatisfactory. Of the remaining terms that are included in both models, all have the same sign, and three of the four are of similar magnitudes. The largest difference is for the coefficient on the number of stations, which is likely because the BART extension to San Francisco International Airport included both additional stations and added service miles. Therefore, leaving service miles out of the regression model means that the stations term picks up the effect to compensate.

The RegARIMA model has the lowest RMSE of the three. Also, the Box-Pierce tests show that there is significant residual autocorrelation in the regression model, risking spurious regression. This is shown by the p-value very close to zero for the regression model. The parameter estimates should remain unbiased, and it appears that they are given their similarity to the RegARIMA model, but the t-statistics are not correct. This can lead to improper inference and specification errors. Given the variables left out of the regression model, compared to the RegARIMA model, it appears that the regression model does suffer from this problem.

To separate the effects of model specification from the model form, Table 4.20 compares the model estimation results for the preferred RegARIMA model to a regression model with the same descriptive variables included. The t-statistics in the

Table 4.19: Three models of BART boardings

Model Characteristics							
Dependent variable	BART boardings						
Date range	Jan 2001 to Mar 2015						
Predictive Variables							
		ARIMA		Regression		RegARIMA	
Description	Lag	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
Auto-regressive coefficient	1	-0.598	-8.54				
Auto-regressive coefficient	2	-0.449	-6.41				
Moving average coefficient	1					-0.5701	-5.08
Moving average coefficient	2					-0.2827	-2.74
Seasonal moving average coefficient	S1	-0.863	-11.42			-0.6603	-8.44
Weekday service miles, 1000s	D(0,12)					2712	2.07
Number of Stations	D(0,12)			7613	2.23	5472	5.18
Employment in 4-county area				0.1827	3.50	0.2027	10.96
Percent of 4-county employment in SF						8099	2.10
Cash fare (2010 $)				-23490	-2.29	-20795	-2.50
Average car fuel cost (2010 $/mile)						86312	2.74
Days with a BART strike				-19690	-28.77	-19010	-20.97
Model Statistics							
RMSE		9508		6733		4923	
Percent RMSE		2.72%		1.94%		1.42%	
Box-Pierce test p-value		0.986		2.368e-09		0.207	

RegARIMA model are generally larger in magnitude than in the regression model. The coefficient values are all of the same sign, but some differ notably in magnitude. The service miles coefficient is half the value in the regression model as in the RegARIMA model, while the number of stations coefficient is larger in the regression model. The other large difference is for the coefficient on the percent of 4-county employment in San Francisco, for which the regression value is about a third of the RegARIMA value.

Given both that the RegARIMA model fits better, and that the regression model suffers from autocorrelated residuals, the RegARIMA model is selected as the preferred model of BART ridership.

4.4.5.3 MUNI versus BART Comparison

Table 4.21 compares the estimation results for the MUNI RegARIMA model (with constrained employment) and the BART RegARIMA model. Because they are estimated from different data sets, measures such as AICc are not directly comparable.

Table 4.20: Comparison of BART regression and RegARIMA models with equivalent specifications

Model Characteristics					
Dependent variable	BART boardings				
Date range	Jan 2001 to Mar 2015				
Predictive Variables					
		Regression		RegARIMA	
Description	Lag	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1			-0.5701	-5.08
Moving average coefficient	2			-0.2827	-2.74
Seasonal moving average coefficient	S1			-0.6603	-8.44
Weekday service miles, 1000s	D(0,12)	1261	0.23	2712	2.07
Number of Stations	D(0,12)	7214	1.81	5472	5.18
Employment in 4-county area		0.1953	3.21	0.2027	10.96
Percent of 4-county employment in SF		2891	0.74	8099	2.10
Cash fare (2010 $)		-18150	-1.68	-20795	-2.50
Average car fuel cost (2010 $/mile)		77120	1.57	86312	2.74
Days with a BART strike		-19690	-28.74	-19010	-20.97
Model Statistics					
RMSE		6511		4923	
Percent RMSE		1.86%		1.42%	
Box-Pierce test p-value		6.958e-11		0.207	

However, the difference in percent RMSE shows that the BART model fits better against BART data than the MUNI model fits against MUNI data. In addition, the BART model includes more variables of stronger significance. Overall, the BART model inspires more confidence, with the key difference being that it is estimated from a longer time series, so the estimation process is better able to pick up meaningful relationships.

Table 4.21: Comparison of MUNI and BART RegARIMA Models

Model Characteristics					
Dependent variable		MUNI boardings		BART boardings	
Type		$ARIMA(0,1,1)(0,1,0)_{12}$		$ARIMA(0,1,2)(0,1,1)_{12}$	
Date range		Jun 2009 to Nov 2013		Jan 2001 to Mar 2015	
Predictive Variables					
		MUNI		BART	
Description	Lag	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1	-0.3092	-1.67	-0.5701	-5.08
Moving average coefficient	2			-0.2827	-2.74
Seasonal moving average coefficient	S1			-0.6603	-8.44
Weekday service miles, 1000s		7971	2.57		
Weekday service miles, 1000s	D(0,12)			2712	2.07
Number of Stations	D(0,12)			5472	5.18
Weekday service miles on MUNI rail, 1000s		-2777	-1.12		
Average bus runspeed		49853	1.94		
Employment in San Francisco		0.876	fixed		
Employment in 4-county area				0.2027	10.96
Percent of 4-county employment in SF				8099	2.10
Cash fare (2010 $)				-20795	-2.50
Average car fuel cost (2010 $/mile)				86312	2.74
Days with a BART strike				-19010	-20.97
Model Statistics					
Log likelihood		-453.56		-1,571.65	
AIC		917.12		3,165.30	
AICc		918.83		3,167.04	
RMSE		15,401		4,923	
Percent RMSE		3.09%		1.42%	
Box-Pierce test p-value		0.873		0.2068	

4.5 Conclusions

To understand the contributors to ridership changes on MUNI and BART, detailed estimation data were assembled from multiple sources. These include passively collected data such as transit AVL/APC data and faregate counts, and data sources that have emerged over the past several years, such as the LODES and GTFS. A process was developed, as described in Chapter 3 to expand sampled AVL/APC data to ensure that they are representative of the system as a whole. Series available at different spatial or temporal resolutions were combined as appropriate, to achieve a consolidated estimate of key variables. Collectively, this data development process has provided a rich set of variables for consideration in the model estimations, including measures of transit service, reliability, crowding, employment by industry, workers at their residential location, fuel costs, and reported estimates of commute mode shares.

These data were used to estimate time series models of MUNI and BART ridership. All models were estimated based on a second difference, with one of the differences being seasonal. This ensured that the data were stationary, and provided a parsimonious mechanism for including seasonality. Three types of models were considered: ARIMA models, regression models, and regression models with ARIMA errors (RegARIMA). In general, the RegARIMA models provided the best fit and allowed the highest number of logical descriptive variables to be included. The RegARIMA models also avoided a problem of residual autocorrelation observed in the BART regression models.

The unconstrained MUNI estimation produced models that were more sensitive to employment than would be expected given other supporting evidence. For this reason, a second set of models is estimated in which the employment coefficient is constrained to the average number of MUNI boardings per employee. This constrained model is used for the remaining analysis.

It is recognised that a key limitation of the analysis is that it was found to be challenging to estimate significant parameters, and to separate different effects that may move in the same direction at the same time. The challenges in estimation appear two-fold. First, the models are estimated from a relatively short time series because those are the data that are available. In this way, the BART models give somewhat more satisfactory results because they are estimated from a much longer time series, and therefore trends that may show up to some degree over a short period can be confirmed with more data. The second challenge is inherent to the nature of the approach. Because the data only change in one dimension, time, there is a risk of co-linearity in dependent variables, which can make it more difficult to parse out the relative effects of different factors that we believe to be important.

Therefore, it may be valuable to estimate the models instead as panel data models, where the boardings or alightings at each stop or station, in each month, are an observation. This would require additional data development to ensure that the descriptive variables are appropriately associated with the stops, and issues of overlap are properly handled. The structure would be similar to a transit direct-demand model, such as [206, 207, 208, 209], although each of those examples is a cross-sectional model. Only one example of a transit direct demand model could be found using panel data [209], so there appears to be room for further development in the area.

Chapter 5

Understanding Ridership Trends

In Chapter 4, time series models were estimated of ridership on two San Francisco Bay Area transit systems: the San Francisco Municipal Railway (MUNI) bus system and the Bay Area Rapid Transit (BART) system. Those models account for service changes, for other drivers of demand (such as employment changes), and for underlying trends beyond what can be explained by the included variables. In this chapter, those models are applied to understand why BART ridership growth is robust, while MUNI ridership stagnates.

5.1 Introduction

In the previous chapter, Section 4.1 provided background on the research context, providing a basis for understanding how this work fits with past studies examining the drivers of transit demand. This section introduces the planning context in the San Francisco Bay Area, with a focus on the divergent ridership trajectories of these two transit systems. It sets the stage for why the topic is of interest from a transport planning and research perspective.

The 9-county San Francisco Bay area covers approximately 7,000 square miles with a population of 7.5 million residents. It has three core cities of San Francisco, Oakland and San Jose, and features geographic constraints in the Bay itself as well as surrounding mountains. Figure 5.1 shows an overview of the transit system. Of interest in this situation is the MUNI bus system and the BART rapid transit system.

MUNI operates in the city of San Francisco at the northernmost portion of the

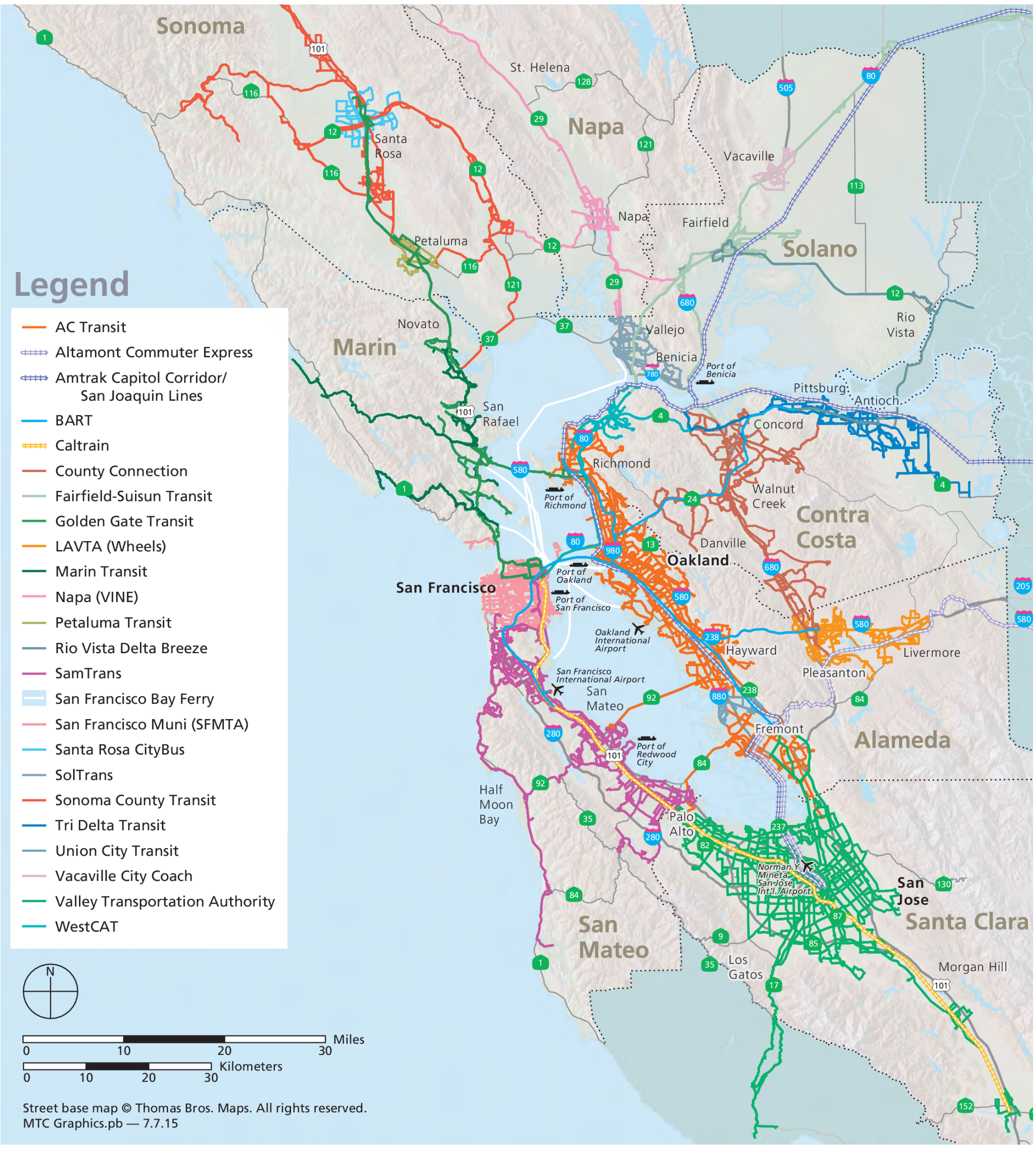


Figure 5.1: Bay Area transit system [142]

peninsula, and specifically serves trips within the city. As of 2014, it carried about 520,000 average weekday riders [142].

BART is a regional rapid transit system that serves four of the Bay Area's nine counties. Importantly, it includes transbay service between Oakland and San Francisco, in competition to the Bay Bridge. Phase one of an extension to San Jose is currently under construction. As of 2014, BART carried about 430,000 average weekday riders [143]. The last major extension to BART was an eight mile extension from Colma to San Francisco International Airport (SFO) that opened in June 2003, although there have been single-station additions since.

Figure 5.2 shows an annotated plot of the ridership trends for both systems, in addition to toll transactions on the Bay Bridge. Over the period from fiscal year 1998 (July 1997 to July 1998) to fiscal year 2013, BART ridership increased by nearly 50%, whereas MUNI bus ridership is nearly identical to its 1998 value.

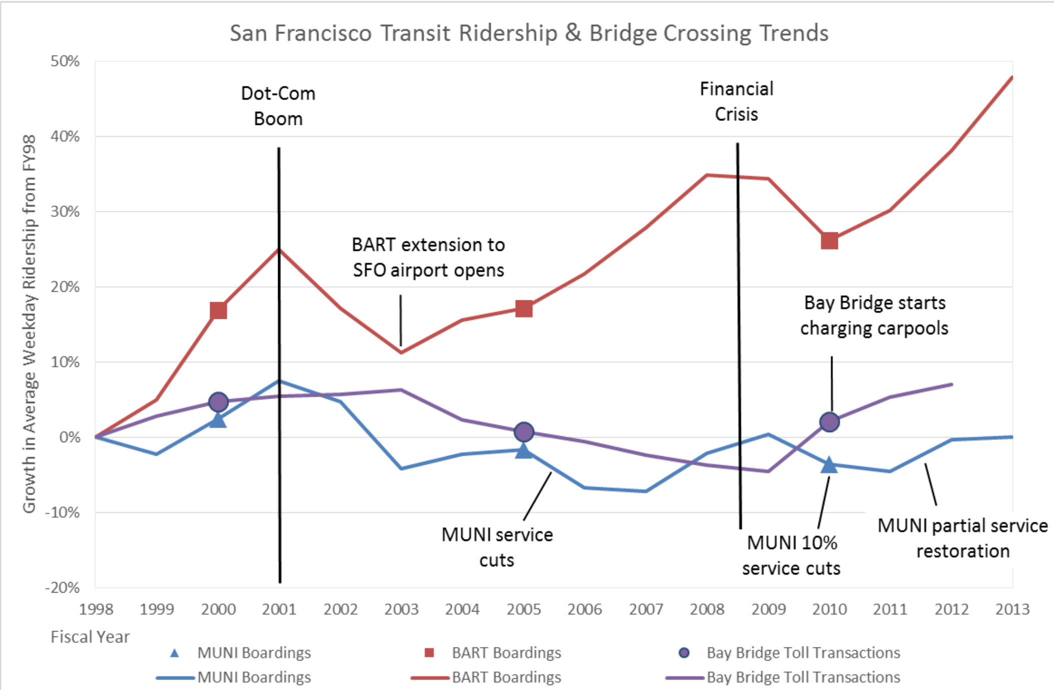


Figure 5.2: Trend in Bay Area transit ridership

Several events are noted on the figure, which appear to have an important effect on the values. The ridership peak in 2001 corresponds to the first dot-com boom, after which the Bay Area lost nearly 20% of its employment in the subsequent bust [210]. The financial crisis in late 2008 appears to have had a similar effect. There is a BART extension to the airport, and several MUNI service changes during this period.

Of particular interest are the very different trends from 2010 to 2013, a period over which more detailed data are available. Over this period, MUNI experiences modest changes in ridership, but BART ridership increases from about 335,000 to 392,000 average weekday riders.

Part of the motivation for this analysis is the author's own experience with the latter. In 2013 I was engaged to provide forecasts to prioritise the seismic retrofit of various BART segments [211]. The issue is that older portions of the BART system do not meet modern earthquake safety standards and need to be rebuilt or retrofitted. The forecasts were used to identify which segments of the system were most critical and should be prioritised to be updated first. This effort used the Metropolitan Transportation Commission's (MTC's) activity-based travel model [212], which had been calibrated to 2010 conditions. When reviewing the initial forecasts, I found that the current 2013 BART ridership was actually higher than the forecast 2020 BART ridership, with the 2020 forecasts including the Phase one extension to San Jose. Upon discussing with the client, some of the transit planners at BART were convinced that this provided evidence of younger generations being more inclined to ride transit, and that all future forecasts should therefore be updated to reflect these changing preferences. While that is one possible explanation, there are other factors that could drive this trend as well.

This analysis uses time series data from both systems to understand what is driving ridership changes on both of these systems, starting from the models estimated in Chapter 4. An important component of that is to identify underlying trends in the data that cannot be sufficiently explained by known and measurable factors, such as service and employment changes.

The remainder of this chapter is structured as follows. The next section presents a review of recent literature that seeks to examine and explain recent trends in travel demand. In Section 5.3, several hypotheses are proposed that may explain the ridership divergence between BART and MUNI. The preferred MUNI and the preferred BART models from Chapter 4 are then applied to understand the contribution of each variable included in the model to the change in ridership. Section 5.4 shows this analysis. In Section 5.5, each of the hypotheses is re-examined in light of the evidence provided by the models. Finding that the explanations in the previous section do not fully explain the trends, additional analysis was performed to consider the possible effects of an ageing population. These are reported in Section 5.6. Finally, conclusions are presented both of relevance to the specific planning questions, and in regards to the approach used to answer those questions.

5.2 Literature Review

This section considers literature that seeks to explain recent trends in travel demand. While the review in Chapter 4 focused on the methods for modelling transit demand, and the variables included in those models, this review focuses more on the explanations, even if those explanations cannot be directly included in the models. This assemblage of explanations seeks to provide a basis for identifying hypotheses that would explain the BART and MUNI divergence, and serve a complement to the quantitative model analysis.

In the US, most transit systems experienced large decreases in ridership from the 1950s through the 1990s, corresponding to the decentralisation of jobs and households, rising incomes and increasing auto ownership. It was in this period of general decline that Kain and Liu [174] studied the factors that drove ridership increases on two successful transit systems, finding that the increases could be explained primarily by large service increases combined with fare reductions during times of employment and population growth. These factors should not be surprising, and show that it is important to first account for the obvious explanations of any differences. In our case, this means examining whether the divergence in ridership between BART and MUNI can be explained by different level-of-service trends or by different land use trends affecting the two systems.

More recently, transit ridership in the US and Canada has increased faster than population growth. Rosenberger et al [213] examine several possible explanations, and suggest that it may be driven largely by growing minority populations and growing income inequality, with minorities and low-income travellers being more likely to ride transit.

The challenges have continued for MUNI, though. In 2005, an influential local non-profit, the San Francisco Bay Area Planning and Urban Research Association (SPUR), published a report about MUNI's "downward spiral" [214]. It observes that MUNI's high cost structure had led to repeated service cuts, which hurt farebox revenue and led to more service cuts. It identified a number of strategies for breaking the spiral, including targeted changes to the route structure focused on improving speed and reliability in the busiest bus corridors. Planning for such changes has been slow, and the budget challenges of the recession were not helpful (see the discussion of service cuts in Section 6.1), but the 5L Fulton Pilot Project is an example of such a targeted change (considered in Section 6.2).

It has been recently observed that auto Vehicle Miles Travelled (VMT) per capita, which had been increasing for decades, has levelled off both in the US and in the UK. A number of studies have examined whether this trend represents a saturation point in the demand for car travel, often referred to as "peak car" [215, 216, 217, 218]. A challenge in answering that question is separating persistent effects from temporary effects, such as recent economic weakness [219]. A selection of studies that seek to explain the relevant trends are discussed next.

Metz [183] observes that per capita travel demand has historically grown with income, but in Britain, this relationship has recently become uncoupled, with personal daily travel ceasing to grow since about 1995. He discusses the possible explanations for such a change. First, he notes that residential development has recently been concentrated in brownfield cites which are more readily served by transit and do not accommodate additional road capacity. This reinforces the need here to consider urban versus suburban growth. Second, he considers the special case of London, which has recently seen robust economic and population growth, corresponding with declining auto mode shares. The net result is that total auto use is flat, as is the road capacity. Interestingly, the introduction of congestion pricing in London is not offered as one of the reasons for declining auto use. This points to a possible consideration of the changing cost of auto travel, either through increased congestion or changing monetary cost. Third, he notes the possible effect of an ageing population, where it is observed that older individuals are likely to have different travel behaviour, but also that there are license holding trends which may be important, where older women are more likely to hold a license than in past cohorts, while younger men are less likely to hold a license than in past cohorts. These factors may not show up as among the key explanatory variables in our models, but could be part of the explanation for trends that the models are not able to fully capture.

LeVine and Jones [184] examine recent travel trends in Britain, focusing on the observation that car travel has recently levelled off in Britain, while train travel has surged. They describe different trends in London, versus the rest of the country. The London findings more relevant to this work due to the parallels between London and San Francisco. Both are economically vibrant cities with strength in the financial and technology sectors, have high levels of road congestion with well-developed transit systems, face imposing housing cost issues, and have diverse populations that may be attracted to the lifestyle of the city. Several major trends they find are:

- While the average car mileage per capita has changed little over the analysis period, there are important differences for certain sub-groups. This points out a limitation to the aggregate analysis used in our study because some of the trends by sub-group may be masked in our aggregate models, but could be considered as a possible explanation for the aspects of the world that our models are not able to capture.

- Changes in the taxation policy of company cars in Britain appear to be related to a drop in car use. Given the different tax structures in the two countries, this change is unlikely to be replicated in the US over our study period.

- There are trends in car mileage by sex and age that may be related to the driver's license holding patterns of the different cohorts. Older women are more likely to hold a license and drive more than in past cohorts, while younger men are less likely to hold a license and drive. While the burden of obtaining a license is notably different in the two countries, this may be a contributing factor.

- The growth in rail travel appears to be related to an increase in the number of people using rail, rather than an increase in mileage for existing customers.

Less information is provided on the causes of these trends, but several possible explanations are offered, including changes to car running costs, reductions in traffic speeds (due to the essentially fixed road capacity in London), investments in improving rail service, income and GDP effects and emerging technology to enable telecommuting or online shopping.

While these two studies are more about car travel, it is interesting to consider whether similar factors may be at play in the stagnation of urban bus ridership in San Francisco. Changes in car running cost or speed could disproportionately affect BART ridership because BART trips tend to be longer than MUNI. Also, with effectively fixed road capacity in San Francisco, the buses will be subject to the same traffic congestion as the cars, potentially causing a shift to rail. If this were the case, we would expect it to be captured by the average bus speed variable in our models. Differences in bus versus rail level-of-service changes may apply. Income is a potentially interesting effect, given that low income travellers are more likely to use transit. Could it be that San Francisco has become so expensive that transit users are priced out? The emergence of new technology certainly could be a factor, given that the adoption of such technology will be uneven across the population, this may be related to the cohort and demographic effects discussed, although one might expect that young men who opt out of driving would choose to live in the city and be more likely to take a bus. It may be possible, though, that they (or other groups) are instead opting for an alternative mode or substituting technology for travel.

The ACS data do track the number of workers who report working at home, so that may offer a rough indication of changes in telecommuting behaviour. Two other possible substitutes are bicycle travel and shared mobility services. Pucher et al [220] reports on a possible bicycle “renaissance” in North America, with strong growth in bicycle mode shares in the US and Canada, but especially in nine large cities, including San Francisco. Brakewood et al [182] found a significant reduction in bus ridership in New York when bike sharing was introduced. In addition, shared mobility modes such as Uber and Lyft are emerging. While they share many characteristics with taxi travel, the technology integration and labour practices of these companies allows them to offer a less expensive and possibly more convenient means of travel to consumers. The Shared Use Mobility Center [221] examines the implications of such service for public transit, and argues that they are more of a complement than competition, due to the high use of shared mobility during off-peak hours, particularly late at night. Together, technology substitution, bicycling and shared mobility may play a role in explaining the trends observed in this research, although with limited data it may be difficult to draw strong conclusions.

This set of possible explanations found in the literature is used as a starting point to propose several hypotheses for why MUNI ridership is relatively flat while BART ridership grows. These hypotheses are described in the next section.

5.3 Working Hypotheses

Several hypotheses are explored that might explain why MUNI and BART ridership trends diverge. These hypotheses are introduced here, and re-visited in Section 5.5 in light of the evidence provided by the time series models. The analysis focuses in detail on the period from 2009 to 2013 when the most complete data are available. However, trends are considered in a broader time context where available.

5.3.1 Different Level-of-Service Trends

A logical explanation for divergent ridership between the two systems is that the transit agencies have made different decisions about the level-of-service they provide.

Figure 5.3 shows the trend in service miles provided by each operator. To provide the best possible long-term estimate of service miles, these time series combine data from two data sources available at different temporal resolutions. Starting in 2009, the General Transit Feed Specification (GTFS) provides a detailed representation of the transit schedules, with the specific dates for which the schedules were in operation. This is the preferred source where it is available. Prior to 2009, the best available source of schedule information are annual reports for each fiscal year from the Transit Statistical Summaries. Therefore, prior to 2009, the time series are “blockier”. For the period where both are available, the fiscal year data are scaled to match the GTFS, and they pivot from this base beforehand.

The data in Figure 5.3 show that MUNI service is highest between 2001 and 2005, with a step down in 2005, and another step down in 2010. After this, there is an increase in late 2010, and several smaller changes. BART has some changes early in the decade, notably with the opening of the extension to SFO in 2003. Service is highest between 2007 and 2009, with a step down in Summer 2009, about nine months before MUNI’s major service cut. BART service is stable after this.

Figure 5.4 shows the trends in fares for MUNI and BART. The cash fare and the average fare are tracked separately. Both are adjusted to 2010 US dollars, indicated on the figure as “2010$”. The cash fare is based on a compilation of press releases and news articles announcing fare changes. The average fare is based on the ratio of reported farebox revenue to reported ridership for each fiscal year. It is adjusted such that the fiscal year totals match the reported values, but fare increases occur on the dates of the cash fare increases, often 1 September. The difference between the MUNI cash fare and average fare is large, because a high number of MUNI riders use monthly passes, or other discounted fares. BART has a much more limited fare discount scheme, so the values are more similar. BART uses a distance based fare scheme, so the reported values are for an average trip.

For both systems, the fares in inflation-adjusted dollars are higher in 2015 than in 2001. The patterns in interim years are more jagged.

Given that the changes are different, both in service provided and in fares, it is reasonable to expect that differing levels of service would play a role in the divergent ridership trends on the two systems. The lumpiness of these plots makes

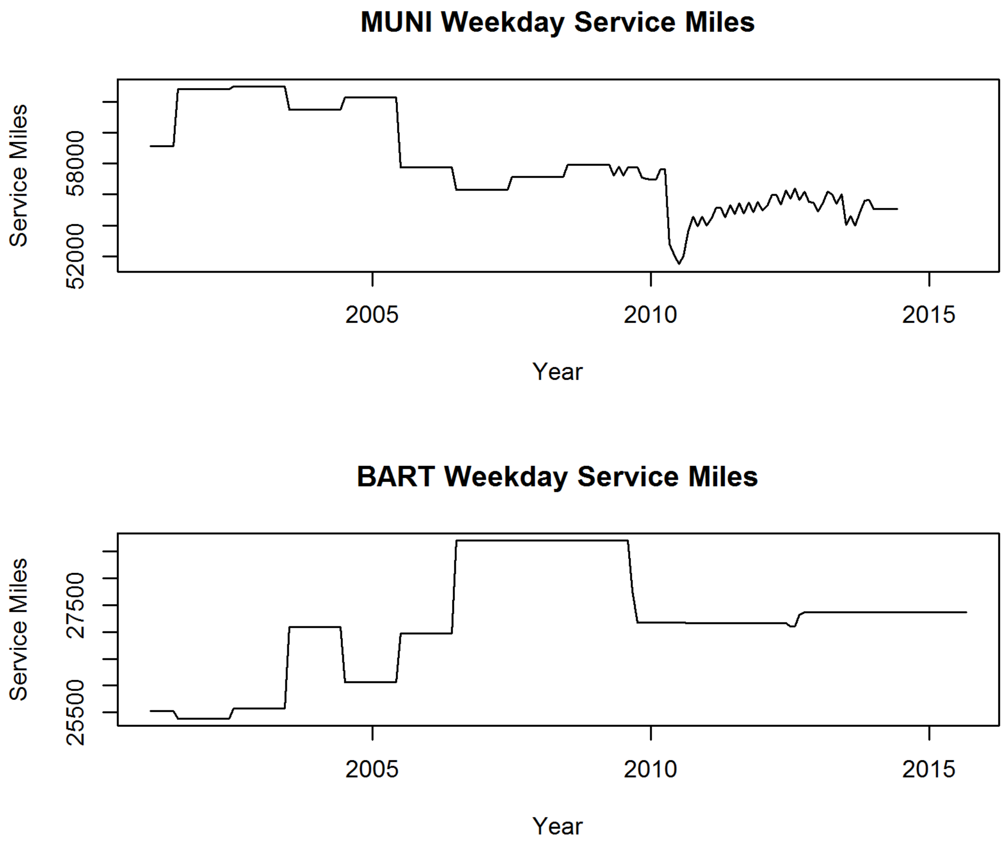


Figure 5.3: MUNI and BART service miles

it clear that such conclusions would be sensitive to the precise starting and ending points chosen for the analysis.

5.3.2 Suburban Growth Exceeds Urban Growth

MUNI and BART serve different travel markets, with MUNI operating exclusively within San Francisco County, and BART serving a larger, 4-county region. A core market for BART is commuters coming into San Francisco from the outlying counties, particularly those east of San Francisco Bay. Given these differences, a higher rate of growth in the outlying counties could drive a higher rate of BART ridership growth. As a starting point for understanding whether that may be an important factor in the divergence, the growth trends are compared for San Francisco versus the 4-county area.

Figure 5.5 shows the employment growth in San Francisco, in the 4-county

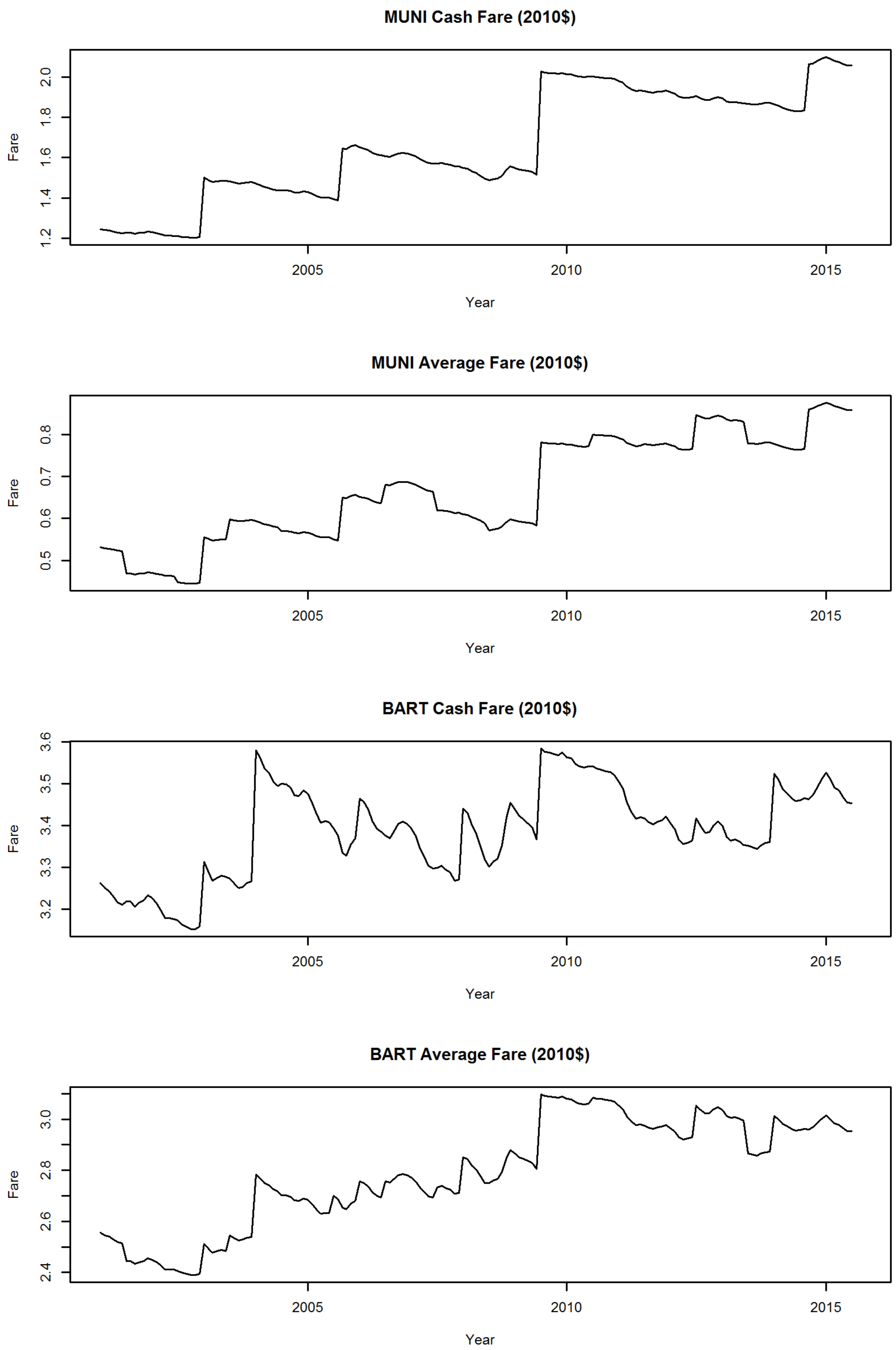


Figure 5.4: MUNI and BART fares

area, and the share of employment in San Francisco. Early in the period, in the wake of the dot-com bust, the share of employment in San Francisco declines. It then steadily increases from about 2006 onwards.

Figure 5.5: Employment in San Francisco and 4-county area

Figure 5.6 shows equivalent plots for population, and Figure 5.7 shows the plots for workers, based on their residential location. The trends in San Francisco and in the 4-county area follow the same basic trends. There is some variation in the share in San Francisco, but the scale of the graph indicates that it is a small variation. The plotted lines are smoother than those in the employment plots these data come from the annual American Community Survey (ACS), and are interpolated to monthly values, whereas employment is measured at a monthly resolution. In additional, there are seasonal employment effects beyond what would be expected for population.

These initial data show a general increase in the concentration of employment in San Francisco, and little change in the concentration of population and workers. These factors are tested in the model estimation, but this initial evidence does not suggest it to be a major driver.

A limitation of this analysis is that employment, population and workers are only considered at a county level. It is possible that variations in the spatial distribution and growth of these at a more detailed level could be an important factor.

5.3.3 MUNI Trips Shift to Rail

It is possible that MUNI bus trips, which are the focus here, are shifting to rail.

To consider this possibility, Figure 5.8 shows the service miles on MUNI bus, rail and cable car, and Figure 5.9 shows the ridership on MUNI bus, rail and cable car, as reported in the annual Transit Statistical Summaries. The service data show high points in the share of bus service miles between 2007 and 2009 and from 2011 onwards. The ridership data show that the share of bus ridership is highest before 2007 and after 2013.

It should be noted that the ridership and operating data provided in the statistical summaries is expected to be less reliable than the bus data based on AVL/APC data. Rail ridership, in particular, is based on infrequent manual counts.

5.3.4 Increased Cost of Car Travel

Another possibility is that increases in the cost of car travel have caused a mode shift to transit. It is expected that such a shift would disproportionately affect BART, which serves longer trips, and where the cost differences would likely be amplified

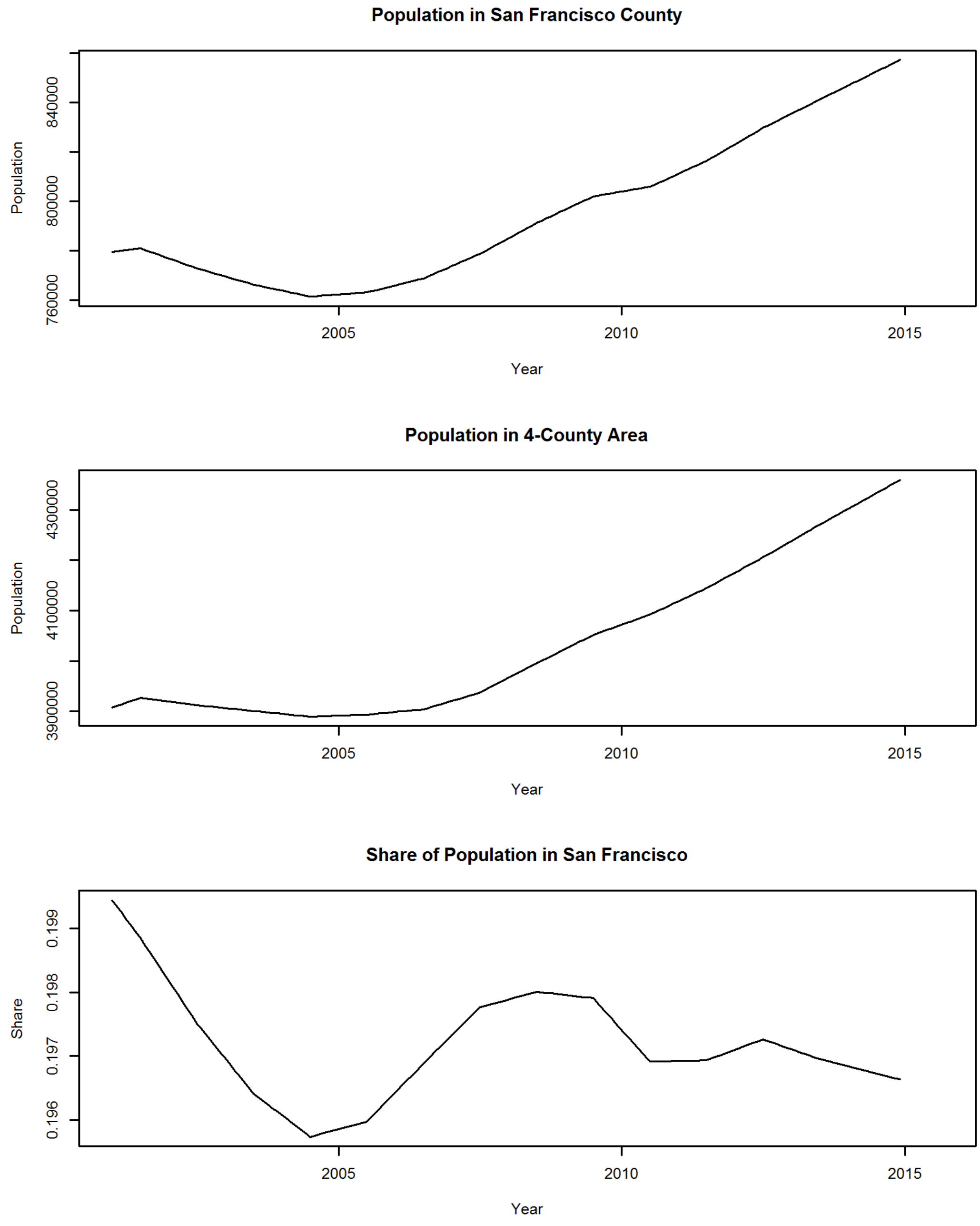


Figure 5.6: Population in San Francisco and 4-county area

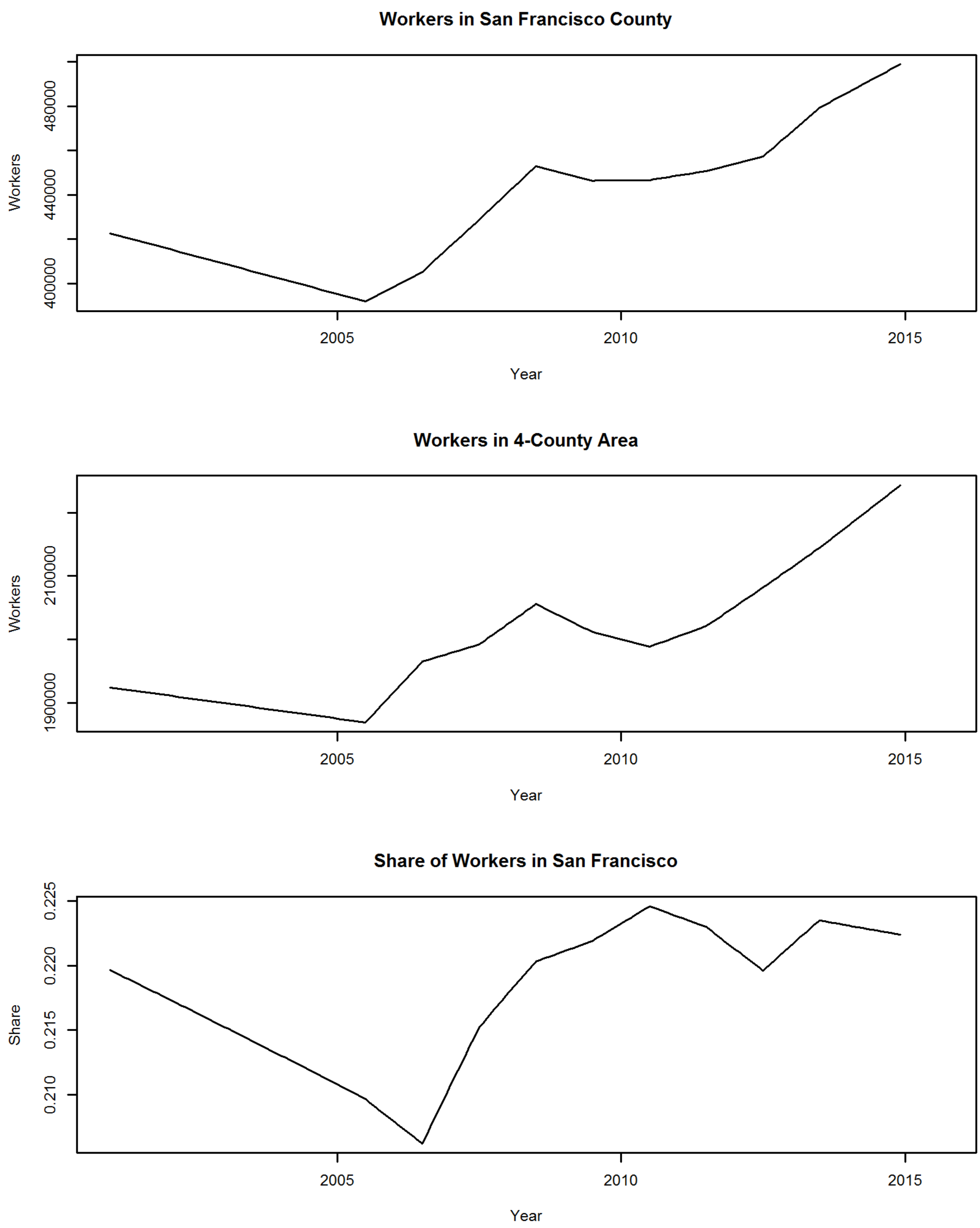


Figure 5.7: Workers in San Francisco and 4-county area

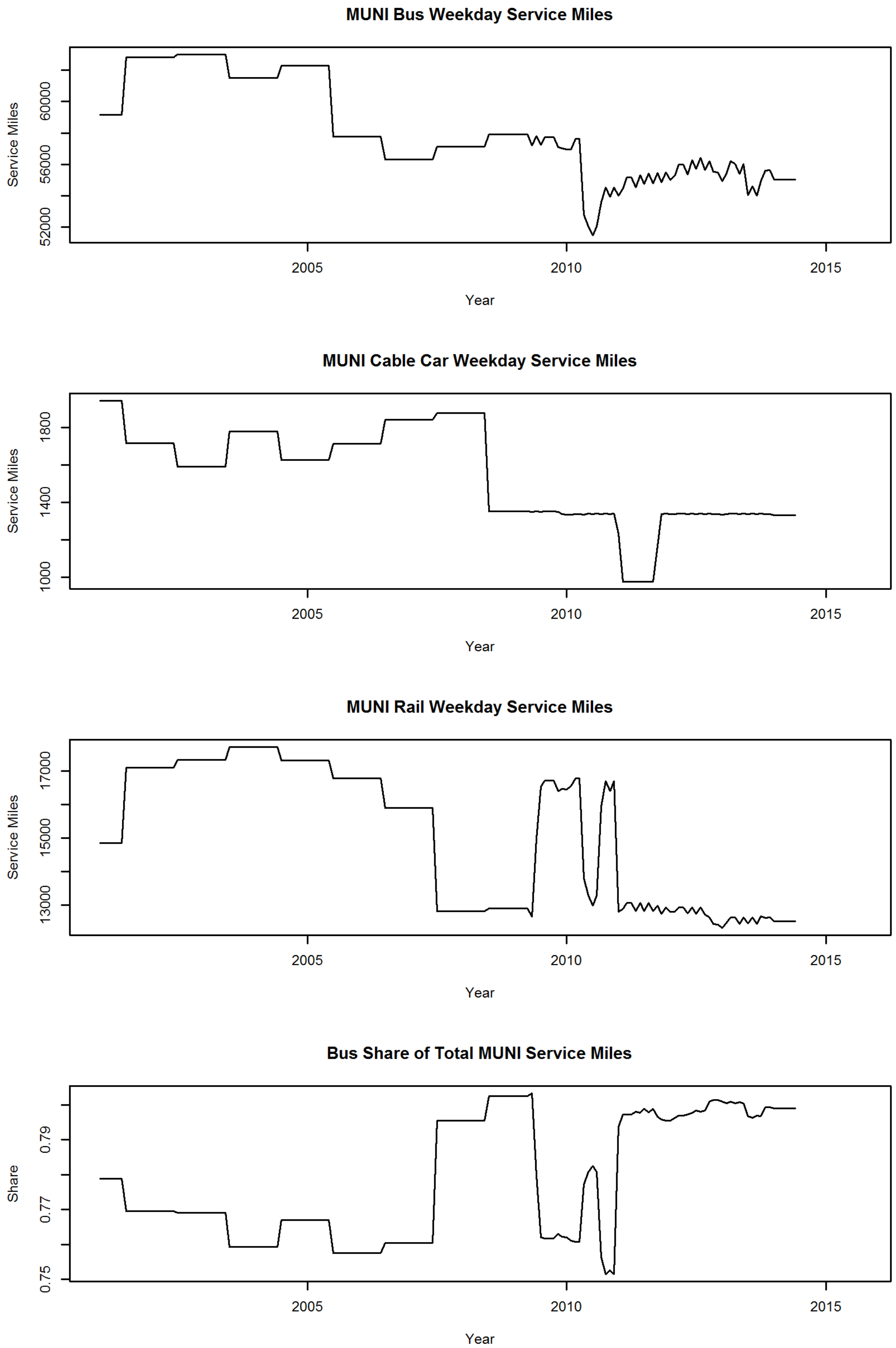


Figure 5.8: Service miles on SFMTA submodes

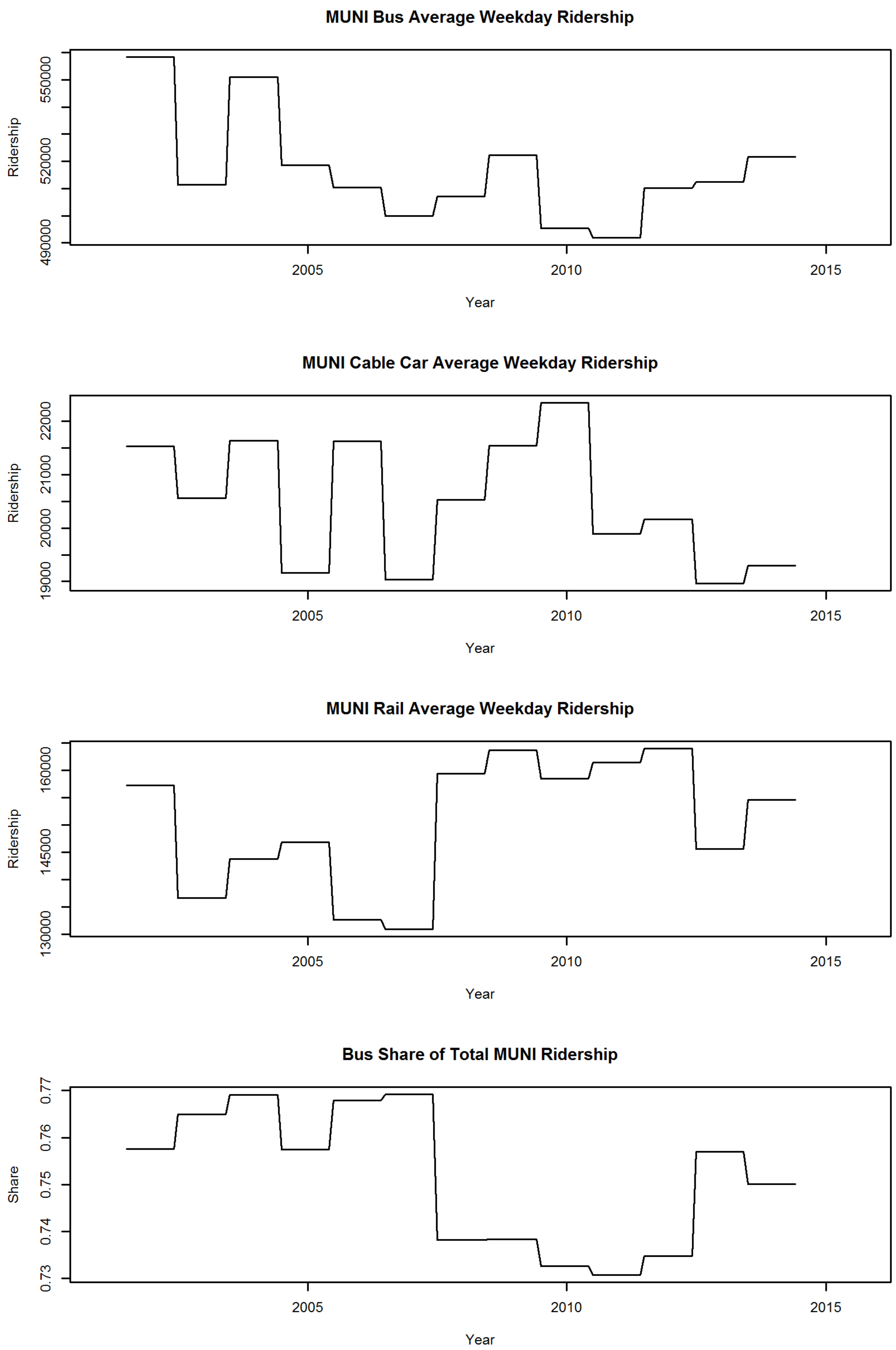


Figure 5.9: Ridership on SFMTA submodes

due to trip length.

Figure 5.10 shows the trends in several measures of car costs. The average fuel cost largely tracks the price of gasoline, but is dampened by improvements to the average fuel efficiency. The average parking costs in the Central Business District are only available for a limited period, and vary from $24 to $28 per day.

Tolls on the Bay Bridge, the competing alternative to BART for crossing the Bay, were increased at several points. Of interest is the increase that occurred in July 2010, when peak tolls were increased, but not off-peak tolls, and a peak carpool toll was introduced for the first time. The goal of this change was to shift trips to the off-peak period and smooth congestion [222, 223].

To test this hypothesis, these terms were included in the time series model estimations. They currently focus on monetary costs as opposed to travel time costs.

In addition, aggregate congestion measures for the Bay Area were tested in model estimation, but the estimated coefficients were found to be statistically insignificant. A separate effort was started to use probe vehicles and a traveller information system to track how car travel times change over this period in a more detailed manner, but those measures are not sufficiently complete to be included in this analysis.

5.3.5 Transit Riders are Priced out of San Francisco

San Francisco has been at the centre of the recent technology boom, with that and other factors shifting the character of the city as it grows. Issues associated with gentrification and inequality have recently been a source of tension [224]. Given that lower income travellers are more likely to use transit, is it possible that transit users are being priced out of San Francisco?

In consideration of this possibility, Figure 5.11 shows the trends in households by income group over this period. These data show that the net growth has been exclusively among households with an annual income of $100,000 or more. Those households double, from about 80,000 to 160,000. All other groups have declined not only in share, but in absolute numbers. To put the values in context, the median annual household income in the United States is about $50,000. A household with

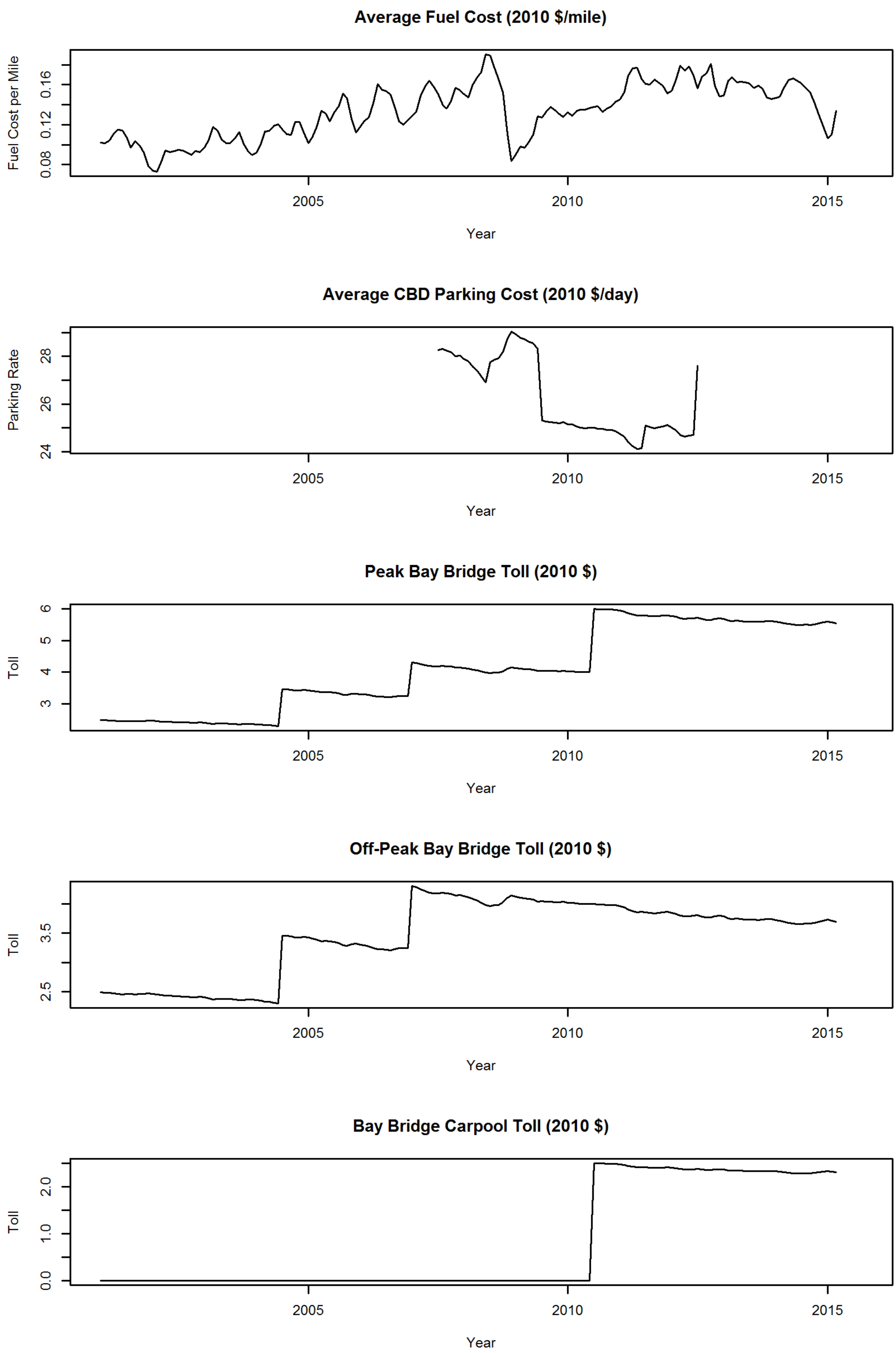


Figure 5.10: Car travel costs

an annual income of $100,000 would be in the top 10% of U.S. households, and the top 25% of Bay Area households.

The income breakpoints used here are aligned with breakpoints in the American Community Survey (ACS), which is the source of the data. Those breakpoints are not adjusted for inflation, so part of the effect is an overall growth in nominal incomes, but the effect is stark enough that there appears to be a true shift in the households as well.

Figure 5.12 examines this trend using the earnings of workers instead of household income, and finds that the growth in workers living in San Francisco is exclusively among workers earning $75,000 or more in annual wages.

Figure 5.13 uses the same grouping of workers, but examines the reported commute mode share of those workers. The data show a strong upward trend in the share of workers earning $75,000 or more who commute by transit, increasing from 22% in 2001 to 34% in 2015. This transition largely closes the gap with the transit mode share of workers earning less than $75,000.

The net effect of changes in households by income, workers by earnings, and work mode shares are not clear. These factors are tested in the model estimations.

5.3.6 The "Uber Effect"

Examining the annual trends in commute mode shares in San Francisco reveals a strong growth in the share of commutes made by walk, and made by taxi, bike and other. Figure 5.14 shows these trends, with the bottom chart indicating that their combined effect is to increase the share from 13% in 2005 to 18% in 2015. For comparison, the transit mode share over this period (from Figure 5.13) ranges from 31% to 34%.

What might be causing these trends? There are two things that come to mind.

First, recent residential development in San Francisco has been concentrated in the eastern portion of the city, particularly in the South of Market (SoMa) area and along Market Street towards the Civic Center. Both areas are within walking distance of the Central Business District, and generally higher density than the western portion of the city, where the neighbourhoods are filled with row houses. It may

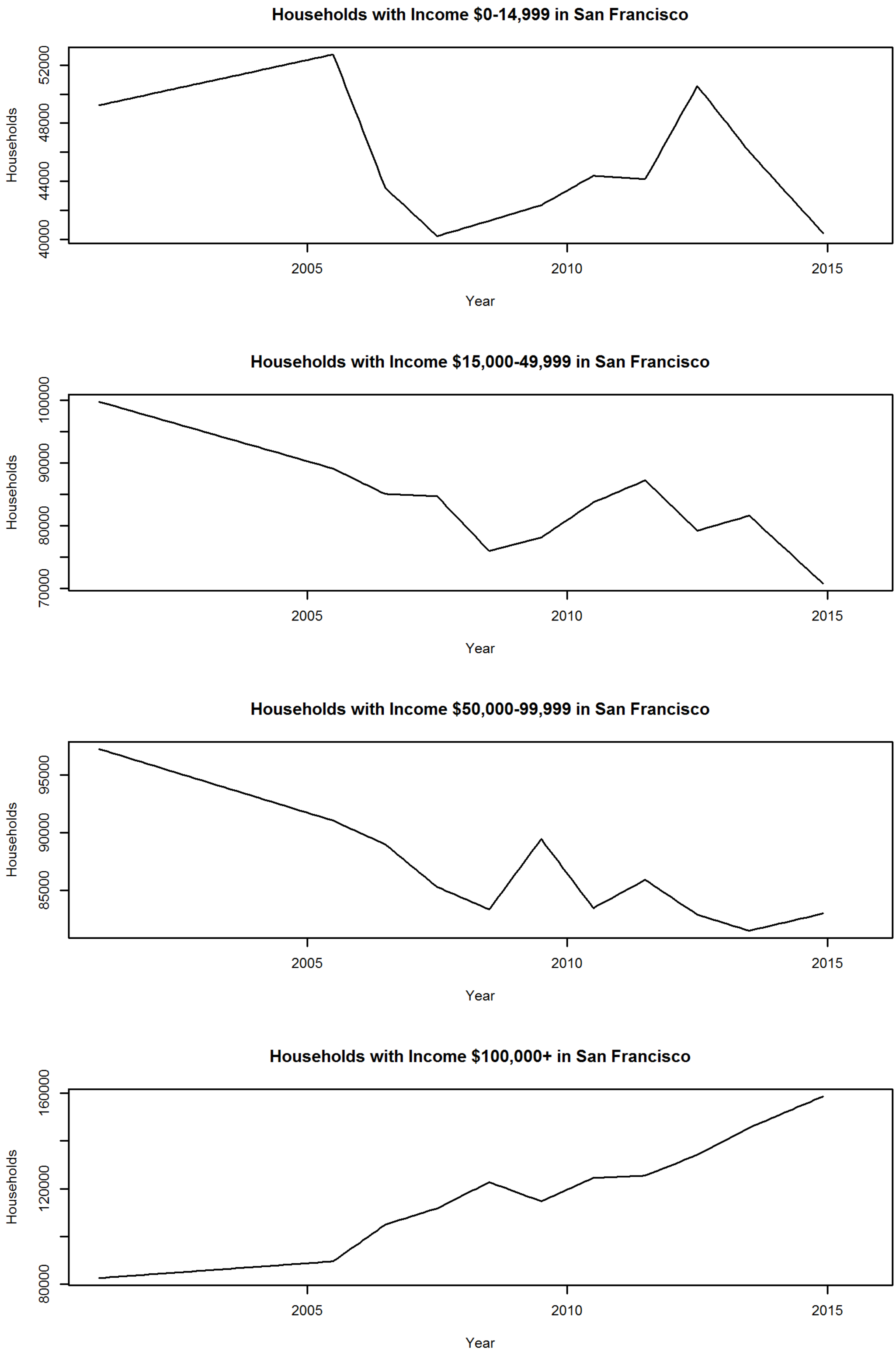


Figure 5.11: Households in San Francisco by income group

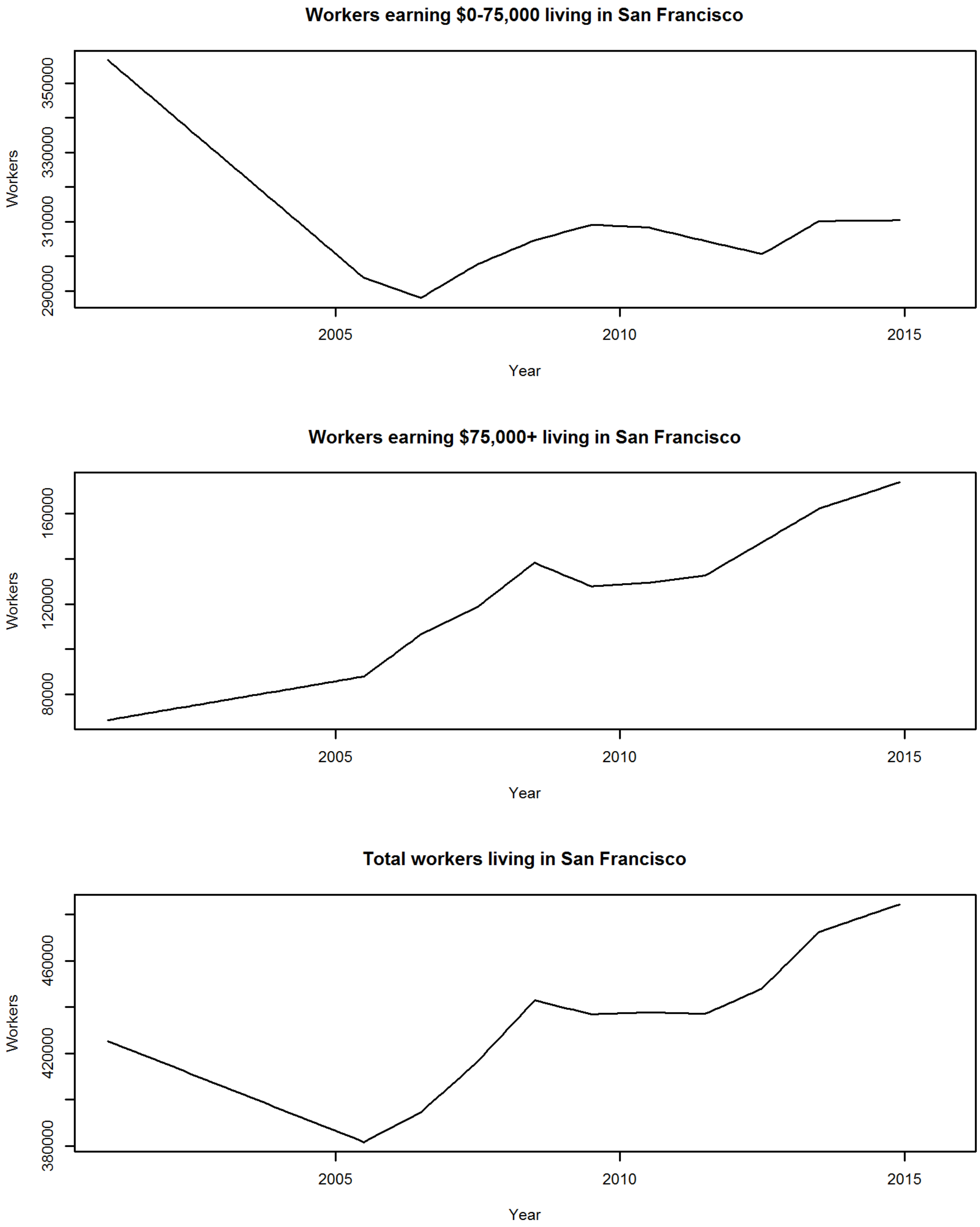


Figure 5.12: Workers living in San Francisco by earnings

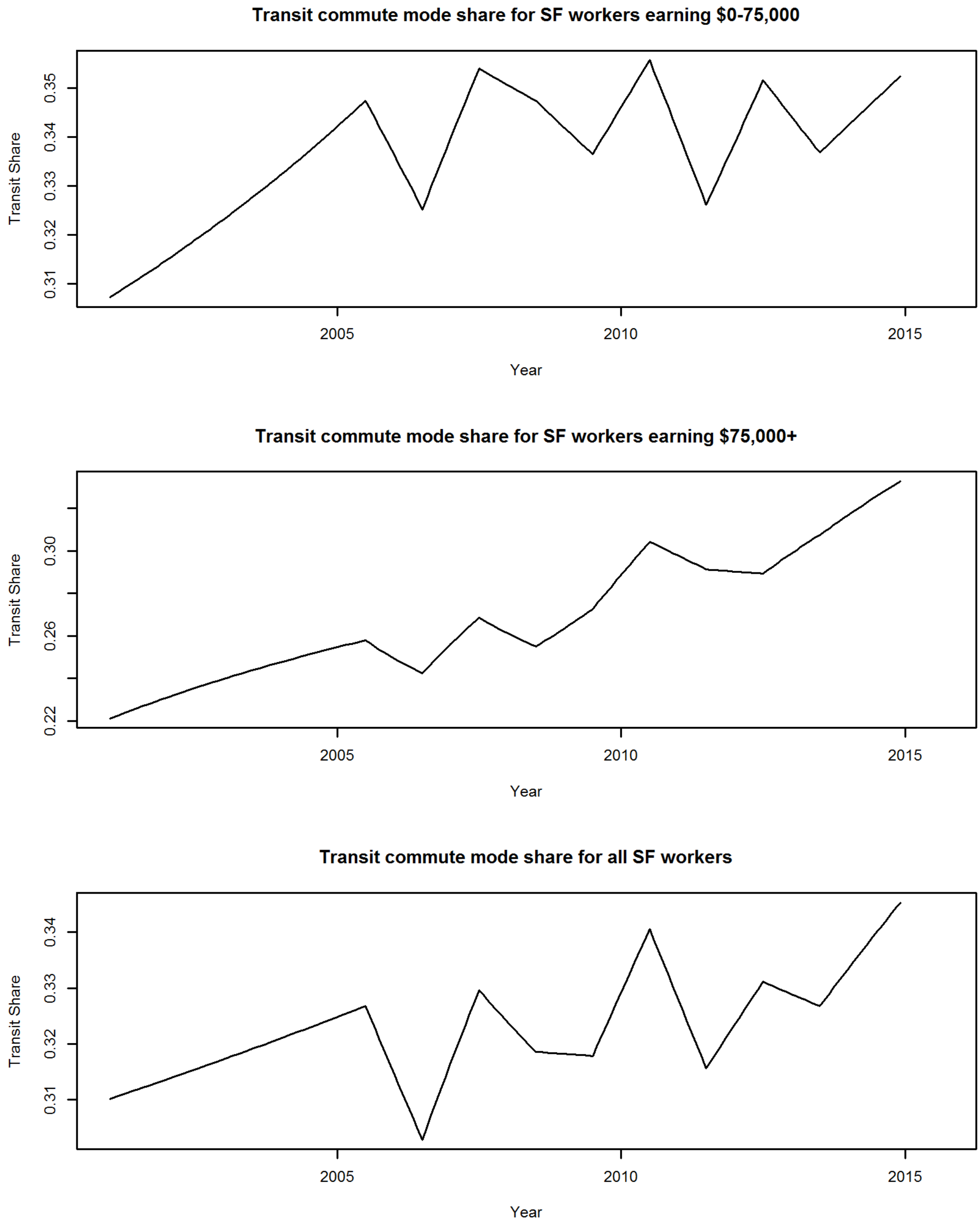


Figure 5.13: Transit mode share by earnings for workers living in San Francisco

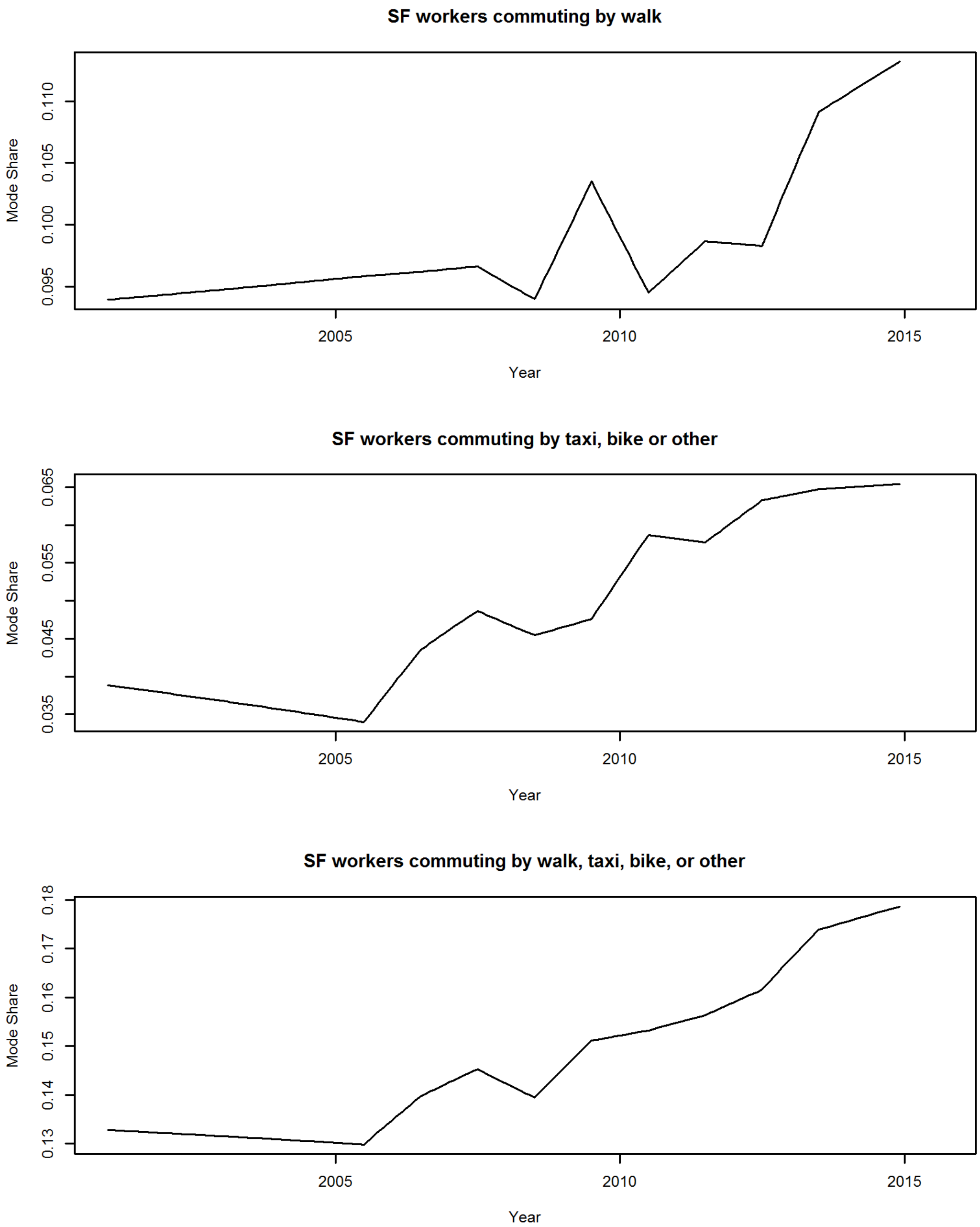


Figure 5.14: Walk, bike, taxi and other commute mode shares for workers living in San Francisco

simply be that the spatial configuration of development is more suited to walking and less suited to riding the bus.

Second, the growth in bike, taxi and other fits with two larger trends that may explain the effect. San Francisco, like other large cities, has experienced an impressive growth of bike travel in recent years. This corresponds both to the advent of bike share systems, and to investments in bicycle lanes and other bike infrastructure [225]. Some research has shown that investments in cycle lanes may have a very small effect on mode shares [226], so it is possible that the growth is driven more by cultural attitudes towards cycling than by infrastructure investments. Regardless, there is a change. In addition, this time period corresponds to the rise of Uber, which was founded in San Francisco in 2009 [227]. Uber, and other shared mobility services, can be both much cheaper than a taxi, and more convenient, particularly for a technology-savvy population, such as San Francisco's. Such trips would be reported as either taxi or other in the ACS. It is reasonable to expect that people who use these modes will use them habitually for many trips, beyond just commuting, and that people who still commute by bus or other modes may make an increasing share of non-work trips by walk, taxi, bike or other.

We hypothesise that there may be an "Uber Effect", where the growth of these alternative modes is directly related to the lack of growth in MUNI ridership. It is logical to expect that such an effect would disproportionately affect MUNI over BART, because the trips are shorter, concentrated in the core city, and do not traverse the Bay or equivalent physical barriers.

The causal mechanism of such an effect could work in several directions. It could be that these alternative modes are becoming more attractive, such as with the large cost savings offered by Uber over taxi. It could be that MUNI is becoming less attractive, and these modes are picking up the slack. Finally, it could be part of a larger demographic or cultural shift in how people consume transportation. This research is insufficient to explain the causal mechanism, but it does explore the relationship between these trends and MUNI and BART ridership to see if the correlations support the idea of an "Uber Effect".

5.4 Understanding Demand Changes

The Regression with ARIMA Errors (RegARIMA) models estimated in Chapter 4 were applied to understand the factors contributing to changes in BART and MUNI ridership. For convenience, the preferred MUNI and BART model estimation results are repeated here in Table 5.1 and Table 5.2. Of particular interest is the divergence of ridership between the two systems during the 2009 to 2013 period, where BART ridership increases and MUNI ridership does not. The goal is to understand why these diverge.

Table 5.1: RegARIMA Models of MUNI boardings with constrained employment term

Model Characteristics				
Dependent variable	MUNI boardings			
Type	$ARIMA(0,1,1)(0,1,0)_{12}$			
Date range	Jun 2009 to Nov 2013			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Moving average coefficient	1	-0.3092	0.1852	-1.67
Weekday service miles, 1000s		7971	3105	2.57
Weekday service miles on MUNI rail, 1000s		-2777	2488	-1.12
Average bus runspeed		49853	25692	1.94
Employment in San Francisco		0.876	fixed	fixed
Model Statistics				
Log likelihood		-453.56		
AIC		917.12		
AICc		918.83		
RMSE		15,401		
Percent RMSE		3.09%		
Box-Pierce test p-value		0.873		

This analysis is done in three parts. First, the model elasticities are calculated and compared to published values, where available. Second, for each variable included in the model, bivariate area plots are presented comparing the actual ridership to what the ridership would have been if that variable remained constant. This provides an indication of how much that variable contributes to the change. Third, tables are presented which show the change in each variable between September 2009 and September 2013, and calculates the ridership change associated changes to each variable.

Table 5.2: RegARIMA models of BART boardings

Model Characteristics				
Dependent variable	BART boardings			
Type	$ARIMA(0,1,2)(0,1,1)_{12}$			
Date range	Jan 2001 to Mar 2015			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Moving average coefficient	1	-0.5701	0.1122	-5.08
Moving average coefficient	2	-0.2827	0.1032	-2.74
Seasonal moving average coefficient	S1	-0.6603	0.0782	-8.44
Weekday service miles, 1000s	D(0,12)	2712	1310	2.07
Number of Stations	D(0,12)	5472	1057	5.18
Employment in 4-county area		0.2027	0.0185	10.96
Percent of 4-county employment in SF		8099	3860	2.10
Cash fare (2010 $)		-20795	8332	-2.50
Average car fuel cost (2010 $/mile)		86312	31504	2.74
Days with a BART strike		-19010	906.5	-20.97
Model Statistics				
Log likelihood		-1571.65		
AIC		3165.30		
AICc		3167.04		
RMSE		4923		
Percent RMSE		1.42%		
Box-Pierce test p-value		0.2068		

5.4.1 Model Elasticities

Table 5.3 shows the elasticity of MUNI ridership with respect to a change in each of the descriptive variables in the RegARIMA model from Table 4.14. The model follows the form of Equation C.26 in Appendix C. For each variable, x, the elasticity is calculated as $\beta\frac{x}{y}$, where y is the demand [228]. Because the elasticity depends on the value of x and y, it changes over the course of the time series. Therefore, the lowest and highest values are presented. The coefficient and mean value of each variable are also shown. For reference, the average MUNI ridership over this period is 501,734.

The elasticity of MUNI ridership with respect to a change in service miles is between 0.89 and 1.05. This means that a 1% increase in service miles would result in ridership increasing between 0.89% and 1.05%. The elasticity for changes in MUNI rail service is between -0.10 and -0.16. It is logical that the value would be negative and smaller in magnitude for changes in service to a different mode than to the same mode. The model is more elastic with respect to changes in bus speed with the values ranging from 1.26 to 1.48.

The elasticity with respect to changes in employment in San Francisco is between 2.19 and 2.90. This elasticity is at the high end of several other studies that have calculated the elasticity of transit ridership with respect to changes in employment. In metropolitan Houston and San Diego, Kain and Liu [174] estimated an elasticity of 0.25. Chen et al. [129] found it to be 0.6 in a study of New Jersey Transit. Gomez-Ibanez [173] estimated an elasticity of 1.25 to 1.75, and Yanmaz-Tuzel and Ozbay [180] estimated a range of 1.6 to 2.7 for New Jersey Transit. Overall, these studies present a wide range of elasticities, with the estimated values at the high end.

Table 5.3: Elasticities from MUNI RegARIMA model with unconstrained employment term

Description	Lag	Coefficient	Mean Value	Elasticity Low	Elasticity High
Weekday service miles, 1000s		8,536	55.69	0.89	1.05
Weekday service miles on MUNI rail, 1000s		-4,352	13.82	-0.10	-0.16
Average bus runspeed		63,927	10.65	1.26	1.48
Employment in San Francisco		2.201	570,931	2.19	2.90

For comparison, Table 5.4 shows the elasticities calculated using the MUNI RegARAMA model with a constrained employment coefficient. As with the previous table, the lowest and highest elasticity from the time series are reported to provide a range of values. The biggest difference is that the elasticity with respect to changes in employment is much lower, between 0.87 and 1.16. The values for rail service and bus runspeed are also somewhat lower. This employment elasticity is more towards the middle of the range of published values, and it is intuitively logical that ridership would increase proportionally to employment, and not faster. For these reasons, the model with a constrained employment term is used for the remaining analysis in subsequent sections.

Table 5.5 shows the elasticity of BART ridership with respect to a change in each of the descriptive variables in the preferred BART RegARIMA model, show-

Table 5.4: Elasticities from MUNI RegARIMA model with constrained employment term

Description	Lag	Coefficient	Mean Value	Elasticity Low	Elasticity High
Weekday service miles, 1000s		7,971	55.69	0.83	0.98
Weekday service miles on MUNI rail, 1000s		-2,777	13.82	-0.07	-0.10
Average bus runspeed		49,853	10.65	0.99	1.16
Employment in San Francisco		0.876	570,931	0.87	1.16

ing the lowest and highest elasticity from the data set. The model follows the form of Equation C.32, and the elasticity is calculated in the same manner as above. The average BART ridership over this period is 344,532.

The elasticity of BART ridership with respect to changes in service miles is between 0.17 and 0.25. This is lower than the equivalent elasticity in the MUNI model, which may indicate that competing modes can be more easily substituted for MUNI trips than for BART trips, given the limited options for transbay travel in particular.

The elasticity for the number of stations is between 0.55 and 0.80.

BART ridership has an elasticity of between 0.95 and 1.38 with respect to changes in employment. The elasticity for the concentration of employment in San Francisco is between 0.58 and 0.82.

The elasticity for BART fares is between -0.16 and -0.25. A fare term could not be estimated for the MUNI model due to the lack of meaningful fare changes during the analysis period.

The elasticity of BART ridership with respect to changes in average fuel cost is between 0.020 and 0.044.

For days with a BART strike, it is between zero and -0.237, although this is not particularly meaningful, given that the number of strike days is always either zero, three or four.

These values generally appear reasonable in comparison to published values [39, 228]. More specific comparisons of changes estimated from these models versus changes estimated from published elasticities are provided in Chapter 6.

Table 5.5: Elasticities from BART RegARIMA model

			Mean	Elasticity	
Description	Lag	Coefficient	Value	Low	High
Weekday service miles, 1000s	D(0,12)	2712	27.12	0.17	0.25
Number of Stations	D(0,12)	5472	42.54	0.55	0.80
Employment in 4-county area		0.2027	1,908,263	0.95	1.38
Percent of 4-county employment in SF		8099	29.53	0.58	0.82
Cash fare (2010 $)		-20795	3.39	-0.16	-0.25
Average car fuel cost (2010 $/mile)		86312	0.13	0.020	0.044
Days with a BART strike		-19010	0.04	0.000	-0.237

5.4.2 Calculating the Factors Contributing to Ridership Changes

The following sections examine the contribution of each factor in the preferred MUNI and BART models to changes in ridership. To visualise the contribution over time, this work takes inspiration from William Playfair's plots of the trade balance between England and Denmark and Norway [229], shown in Figure 5.15. In Playfair's plot, the level of imports is plotted as a yellow line and the level of exports is plotted as a red line, both with time on the horizontal axis. The difference between the two lines is the trade deficit or surplus, and the colour of shading indicates whether there is a deficit or surplus at each time point. This type of plot is generally known as a bivariate area plot, with the graphics in this instance developed using R code developed by Yau [230]. More general discussions of Playfair's contributions are available in [231] and [232].

For this research, the actual time series is plotted against a hypothetical time series that would have occurred if the variable in question remained constant. The difference between the two show the contribution that changes to that variable have made to changes in the time series. Red shading indicates that changes to the variable correspond to a relative decrease in the time series. Green shading indicates that changes to the variable correspond to a relative increase in the time series.

The analysis separately considers each of the descriptive variables in the preferred RegARIMA model. Recall that Equation 4.10 shows the general formula for a RegARIMA model, with Appendix C showing the specific formulas for each

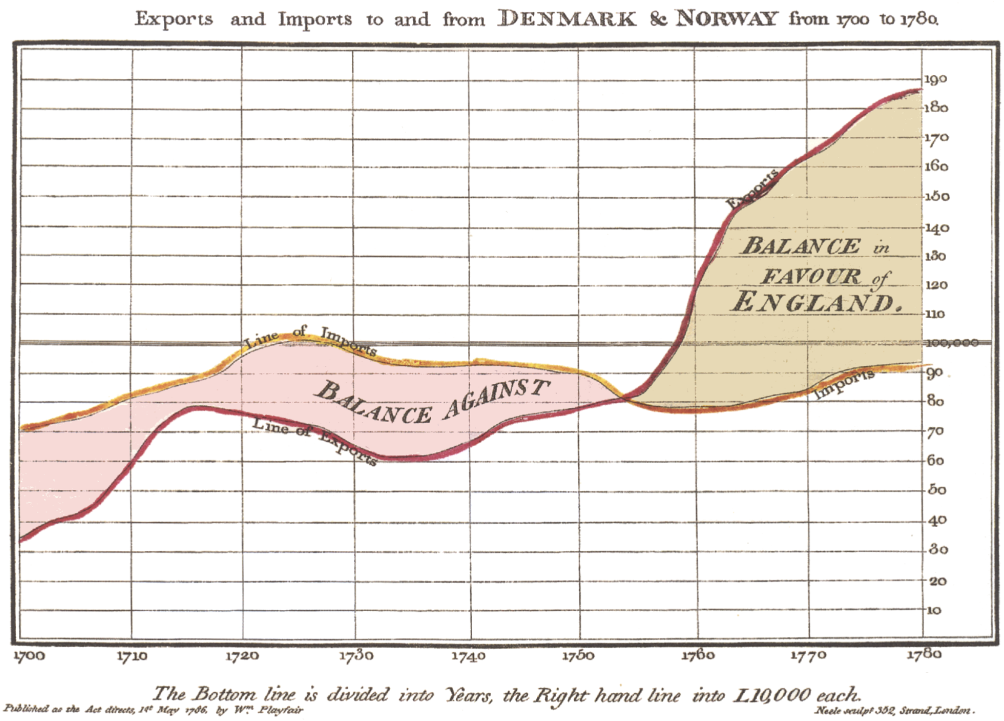


Figure 5.15: William Playfair's plots of English exports and imports to and from Denmark and Norway [229]

model. Starting from the former, the time series with term i held constant is:

$$Y_{i,t} = \beta_i x_{i,0} + \sum_{j=1,j\neq i}^{J} \beta_j x_{j,t} + n_t$$
$$(1-\phi_1 B - \ldots - \phi_p B^p)(1-\Phi_1 B^m - \ldots - \Phi_P B^{m+P})(1-B)^d(1-B^m)^D n_t =$$
$$(1+\theta_1 B + \ldots + \theta_q B^q)(1+\Theta_1 B^m + \ldots + \Theta_Q B^{m+Q-1})e_t \tag{5.1}$$

where $Y_{i,t}$ is the number of MUNI boardings that would occur at time t, if term i is held constant, β_i is the estimated coefficient for term i, $x_{i,0}$ is the value of the variable for term i at time 0, β_j is the estimated coefficient for term j, $x_{j,t}$ is the value of the variable for term j at time t, J is the number of regressors in the model, and all other items are as described previously.

The ARIMA component of the model is the same, and it is only the regression portion that differs. Alternatively, it can be calculated as:

$$Y_{i,t} = y_t - \beta_i(x_{i,t} - x_{i,0}) \tag{5.2}$$

Thus, the difference is based simply on the coefficient times the change in the value of the variable.

In addition to changes that can be explained by each of the regression terms, there will be some portion of the change in the value of the time series that remains unexplained by the variables included in the model. This portion of the change is $n_t - n_0$.

This unexplained portion of the model can be broken into two components: the change in the residual error $e_t - e_0$, and the remaining change $(n_t - e_t) - (n_0 - e_0)$. The change in the residual error can be considered as purely random, but the remaining change accounts for trends and seasonality in the data beyond what can be explained by the regression variables. We refer to this as the unexplained trend. In a way, it is analogous to an alternative specific constant in a mode choice model, because it represents something about the world that the model is otherwise unable to account for. The unexplained trend, and the change due to random error are tracked separately in this analysis to better understand the role they play in understanding changes in demand.

5.4.3 Changes in MUNI Ridership

This section examines the contribution of each factor in the MUNI model to changes in ridership using bivariate area plots, as described above. In all cases, the reference point ($t = 0$) is set to September 2009. Even though this is a few months into the time series, the changes for previous months can still be calculated. September 2009 is selected as the first available month where schools are in session, and to correspond to the differences calculated in Section 5.4.5.

5.4.3.1 Service Miles

Figure 5.16 shows the change in MUNI ridership attributable to changes in service miles. The black line is the observed ridership. The shaded red area shows the difference between the actual ridership, and what the ridership would have been if the service miles remained at their September 2009 levels (assuming the models are correct). If there were no change in service miles, the time series would be shifted to the top of the shaded red/pink area. In this specific plot, there is no green.

This shift is directly related to the trends in service miles discussed in 3.2. The service provided reflects a 10% reduction in May 2010. The cuts are partially restored in September 2010, and there are a series of subsequent, smaller changes.

The plot shows that if it were not for these service cuts, ridership in the summer of 2010 would have been close to 540,000 instead of below 500,000. As late as autumn 2013, ridership remains 25,000 lower than it would have been if not for the cuts to service miles.

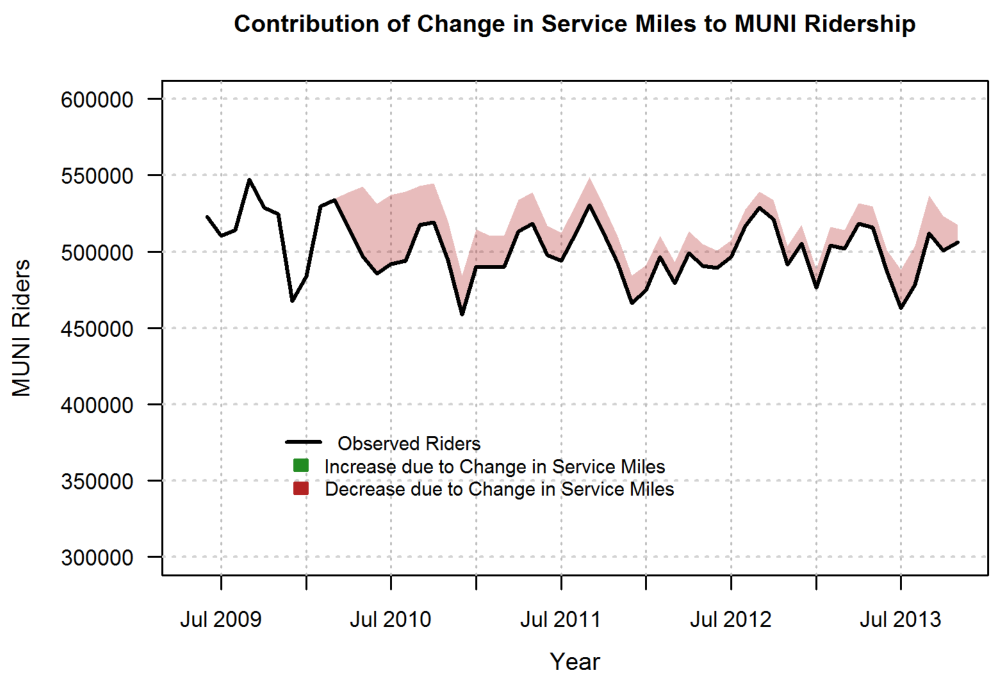


Figure 5.16: Effect of changes in service miles on MUNI ridership, vs. Sep 2009

5.4.3.2 Rail Service Miles

Figure 5.17 shows the change in MUNI ridership attributable to changes in rail service.

The shading in this plot is green. This indicates that cuts to rail service have served to increase the MUNI bus ridership. If rail service were held constant throughout this period, the bus ridership would follow the trend shown at the bottom of the shaded green area, instead of the actual ridership along the black line.

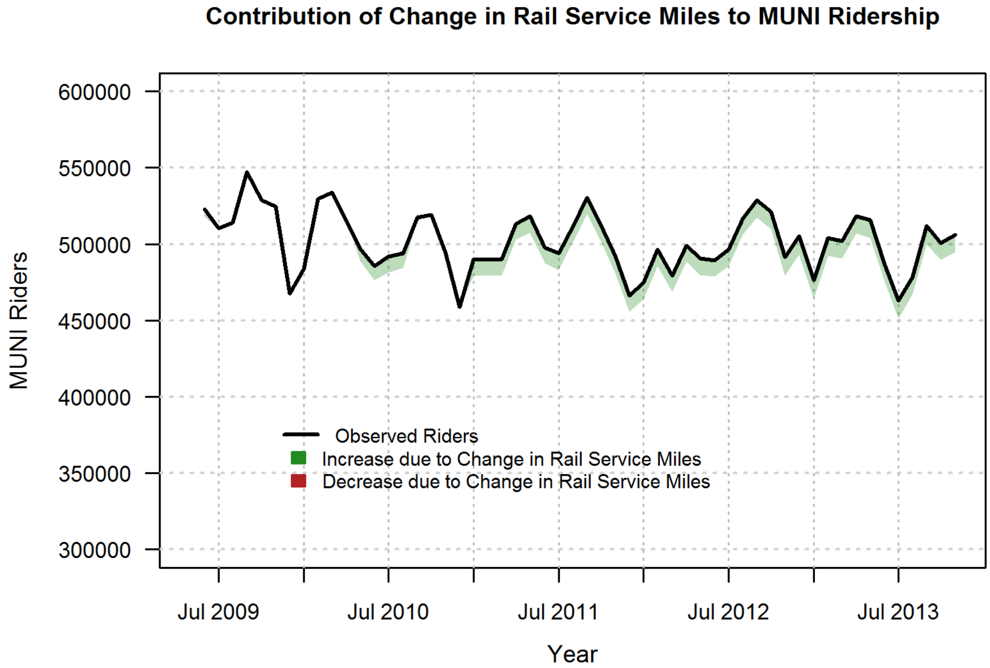


Figure 5.17: Effect of changes in rail service on MUNI ridership, vs. Sep 2009

5.4.3.3 Bus Runspeed

Figure 5.18 shows the change in ridership associated with changes in bus runspeed. The red indicates that slower bus runspeeds correspond to less ridership than would otherwise be expected. It is important to note that the runspeed is not truly independent of ridership. The drop occurs in conjunction with the 2010 service cuts, and it may be that by running fewer buses, there is more crowding that serves to slow the buses down. While there may be some uncertainty in the true nature of the underlying relationship, the basic result is logical, and maintained for this analysis.

5.4.3.4 Employment in San Francisco

Figure 5.19 shows the change in MUNI ridership attributable to the change in employment in San Francisco.

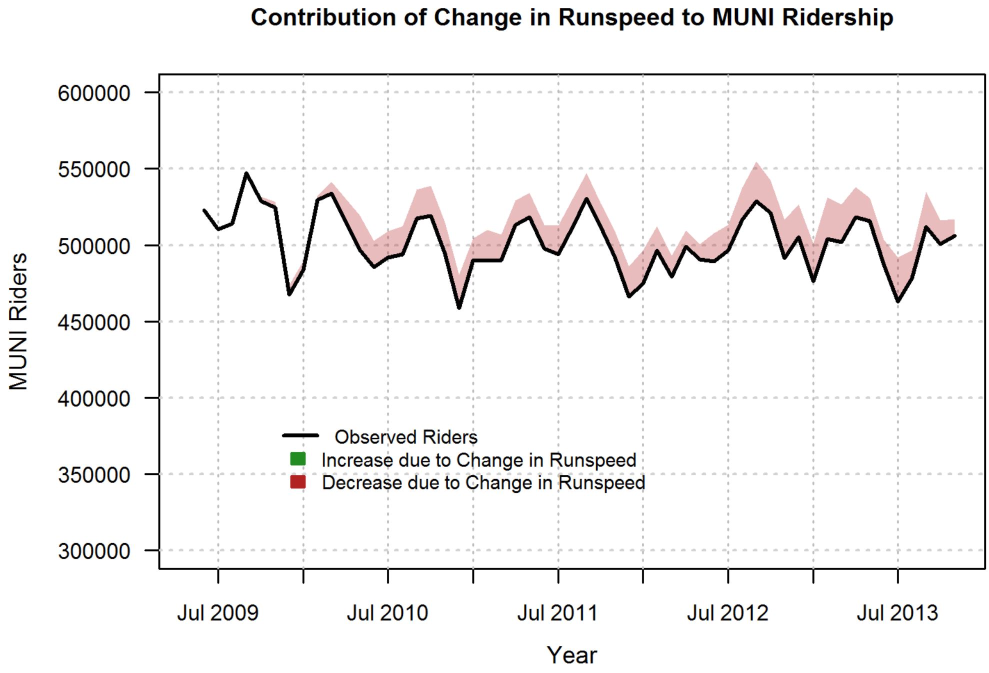


Figure 5.18: Effect of changes in runspeed on MUNI ridership, vs. Sep 2009

The plot shows some small negative effects early in 2009 and early 2010. Starting in the second half of 2010, employment starts to grow and contributes to a relative increase in ridership. The latter portion of the time series in particular corresponds to a period of strong economic growth.

5.4.3.5 Unexplained Trend

Figure 5.20 shows the unexplained trend from the MUNI RegARIMA model. This represents the non-residual change that cannot be explained by the descriptive variables in the regression model. For most of the time span, the trend serves as a substantial drag on MUNI ridership.

5.4.3.6 Residual Error

Figure 5.21 shows the residual error from the MUNI RegARIMA model. This can be interpreted as a purely random component of the ridership. Some values are positive, and some are negative, and there is no clear pattern.

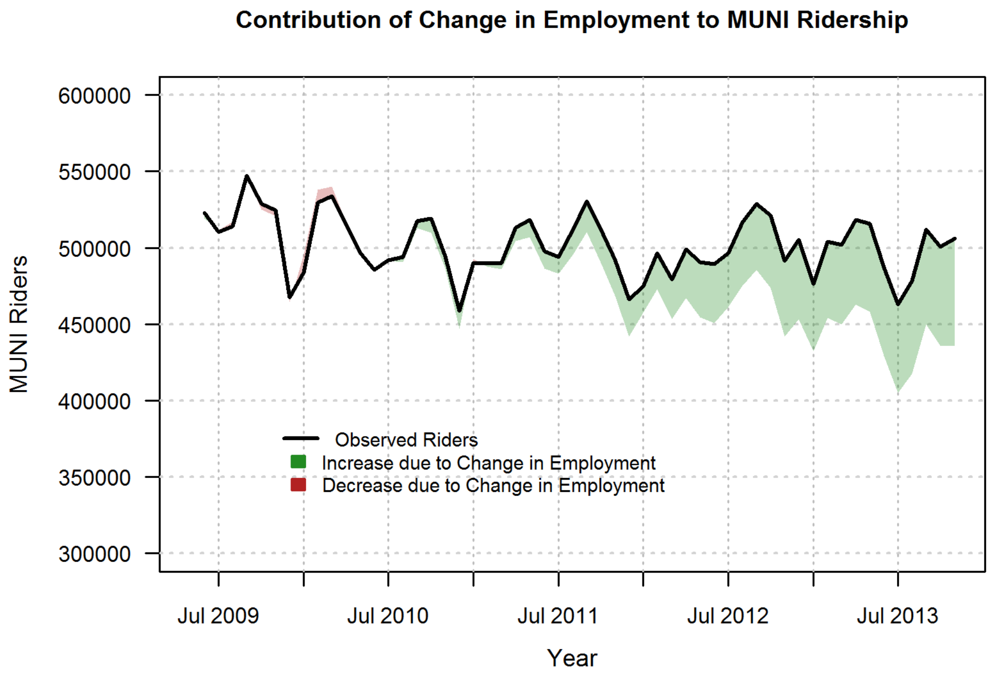


Figure 5.19: Effect of changes in employment on MUNI ridership, vs. Sep 2009

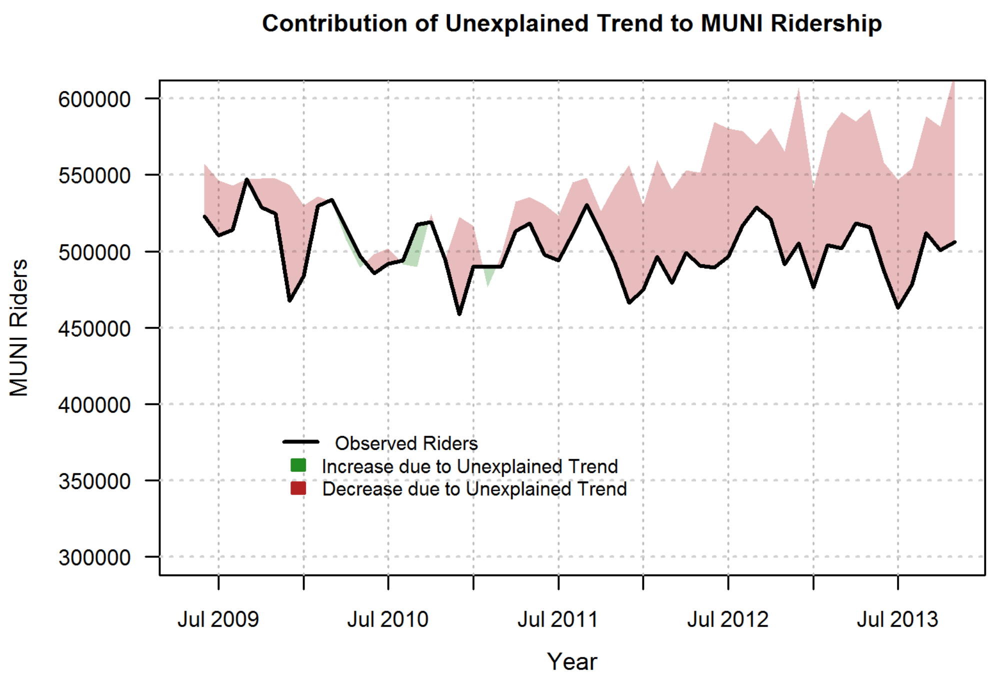


Figure 5.20: Effect of unexplained trend on MUNI ridership, vs. Sep 2009

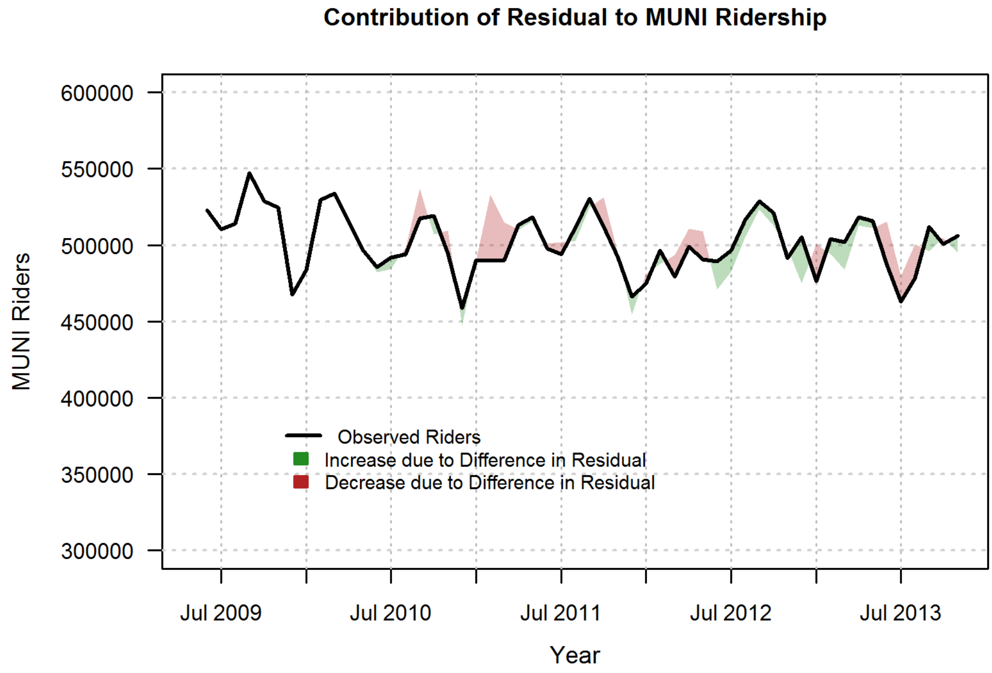


Figure 5.21: Effect of residuals on MUNI ridership, vs. Sep 2009

5.4.4 Changes in BART Ridership

In this section, the same method is applied to examine the contribution of each factor in the BART model to changes in ridership. A longer time series is available for BART. It is desirable both to understand the trends in the full time series, but it is also desirable to be directly comparable to the MUNI plots for understanding the 2009 to 2013 ridership trends in detail. Therefore, for each variable, two sets of plots are shown. The first shows the full time series and sets the reference point to January 2001. The second shows only the period for which MUNI data are also available, and sets the reference point to September 2009.

5.4.4.1 Service Miles

Figure 5.22 shows the contribution of the change in service miles to BART ridership. There are some modest ridership increases associated with service improvements over January 2001 levels starting in about 2004.

Figure 5.23 shows the contribution of the change in service miles to BART ridership for the 2009 through 2013 period. There are some modest decreases due

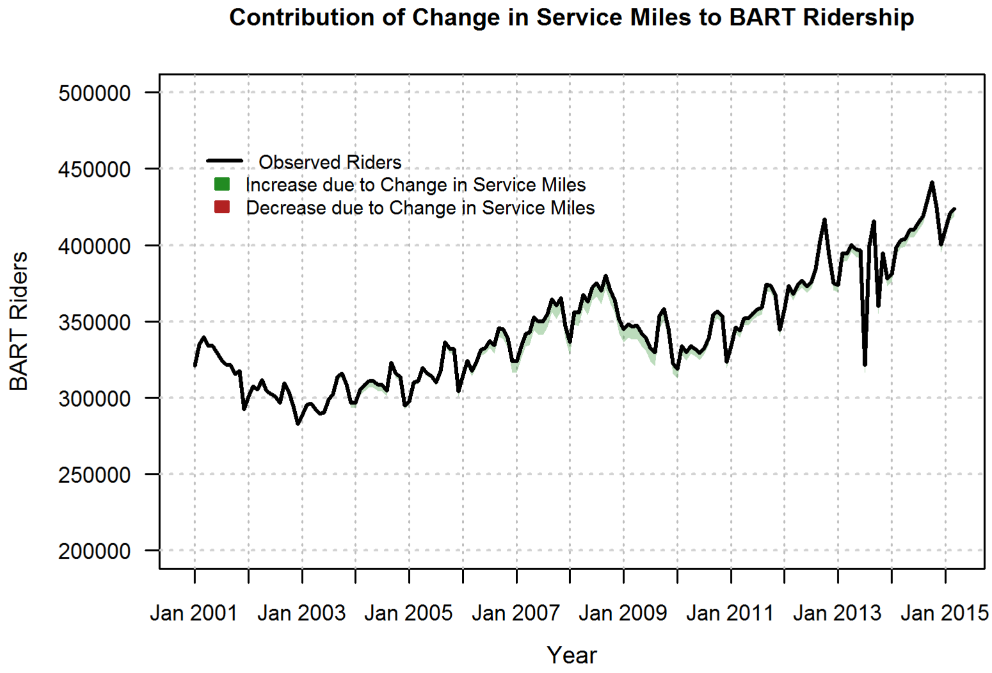


Figure 5.22: Effect of changes in service miles on BART ridership, vs. Jan 2001

to service cuts over this period, relative to the September 2009 reference point.

5.4.4.2 Number of Stations

Figure 5.24 shows the contribution of the change in the number of stations to BART ridership. The biggest change follows the BART extension to SFO, which opened in 2003. Due to the lagged effect on ridership, the full ridership shift does not occur until 2004, after which it persists through the remainder of the period.

Figure 5.25 shows the contribution of the change in the number of stations to BART ridership for the 2009 to 2013 period. The only new station to open in this period is an infill station at West Dublin/Pleasanton, and the effect can be seen in the plot as a modest upward shift.

5.4.4.3 Employment in 4-County Area

Figure 5.26 shows the effect of employment changes on BART ridership. These changes are relative to January 2001 levels.

January 2001 was at the end of the dot-com boom, so employment was at a

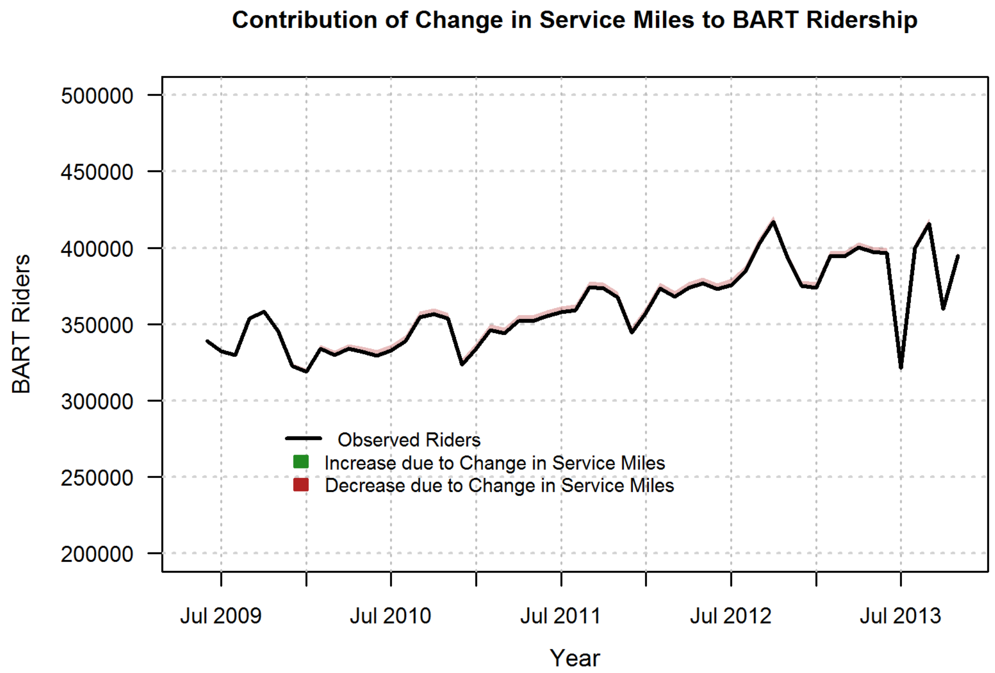


Figure 5.23: Effect of changes in service miles on BART ridership, vs. Sep 2009

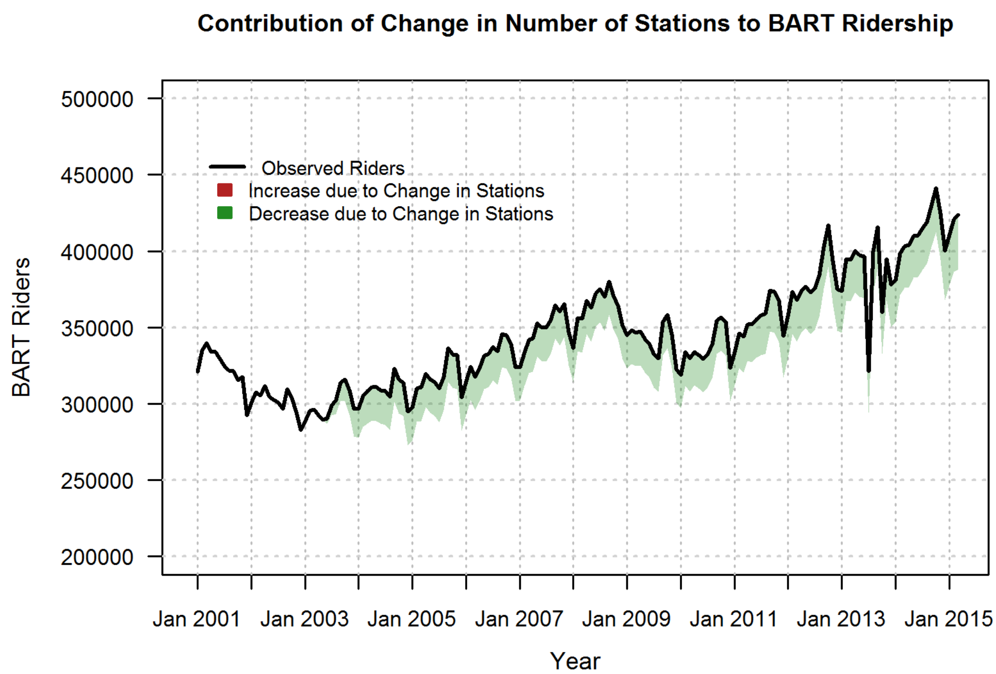


Figure 5.24: Effect of changes in stations on BART ridership, vs. Jan 2001

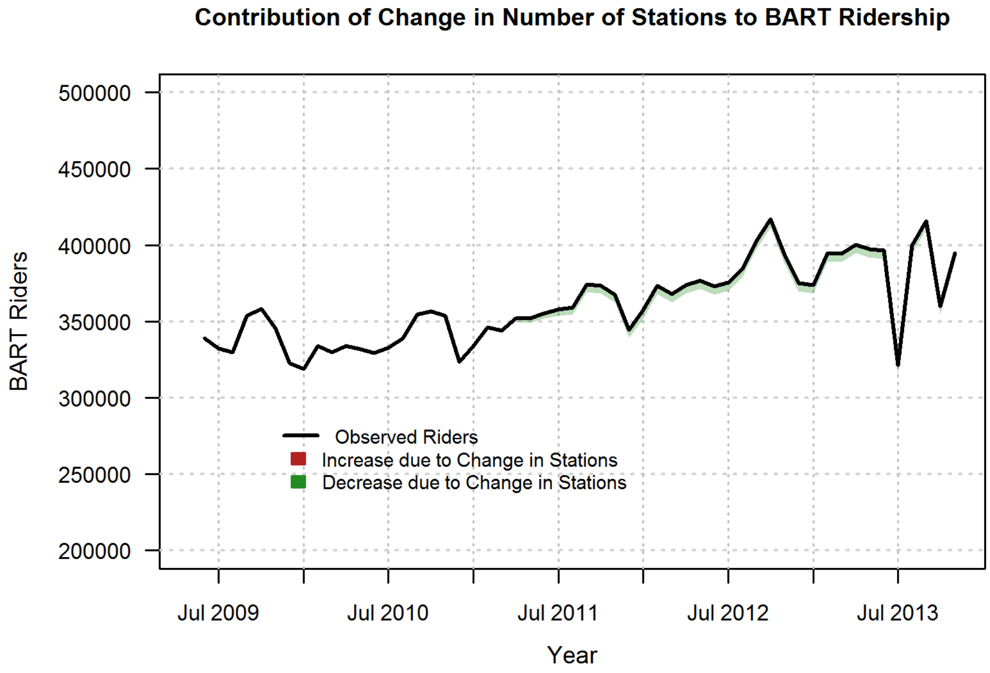


Figure 5.25: Effect of changes in stations on BART ridership, vs. Sep 2009

high point. Employment does not exceed this high point again until 13 years later, where there is some green observed at the end of the plot. The red shading shows the degree to which the dot-com bust, and the recession following the financial crisis pushed BART ridership downwards. If it were not for these recessions, BART ridership would follow a much more steady upwards trajectory.

Figure 5.27 focuses on the 2009 to 2013 period, and shows the change in BART ridership attributable to the change in employment from September 2009. Employment changes make an important contribution to ridership growth over the second half of this period.

5.4.4.4 Percent of 4-County Employment in San Francisco

The percent of 4-county employment in San Francisco is a measure of employment concentration in the region's central city. Figure 5.28 shows the effect of changes in employment concentration on BART ridership.

Over the first half of the period, employment becomes less concentrated, contributing to a relative decline in ridership. Over the second half of the period, em-

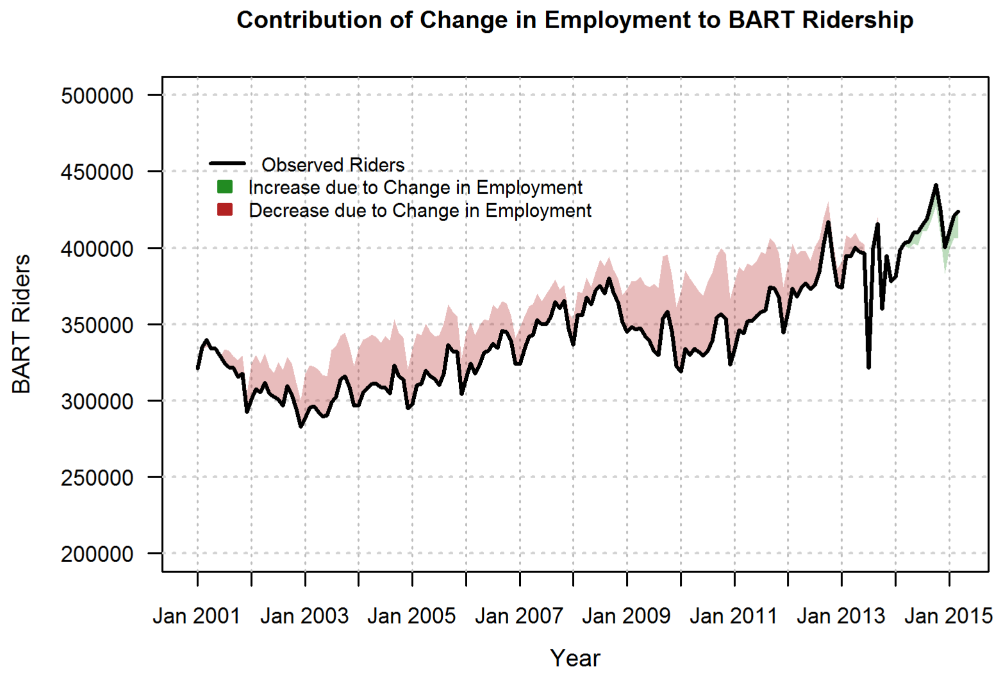


Figure 5.26: Effect of changes in employment on BART ridership, vs. Jan 2001

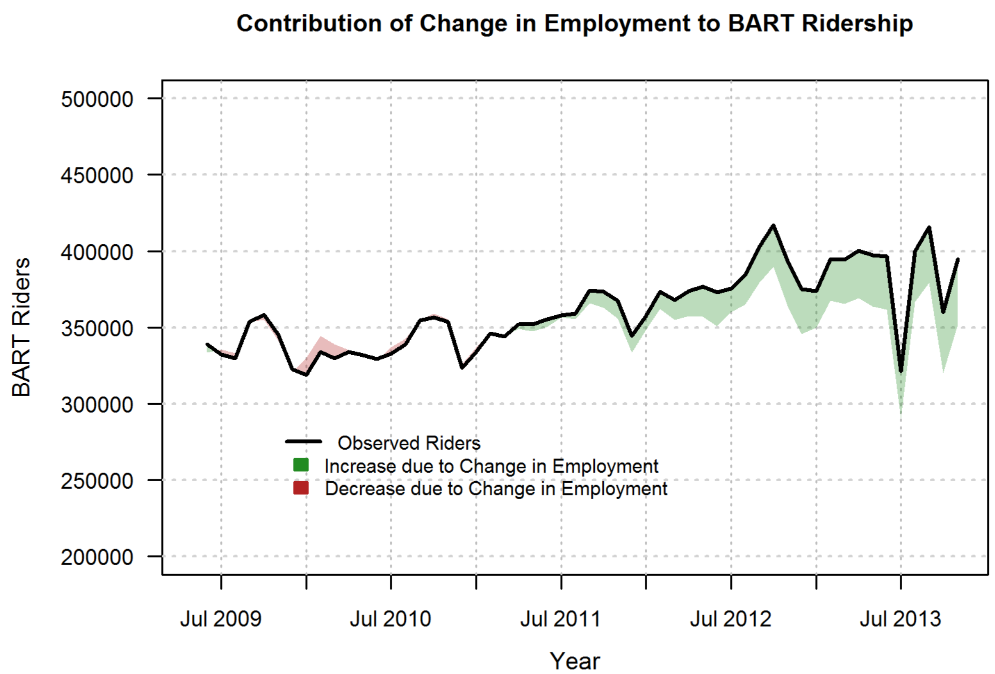


Figure 5.27: Effect of changes in employment on BART ridership, vs. Sep 2009

ployment becomes more concentrated, contributing to a relative increase in ridership.

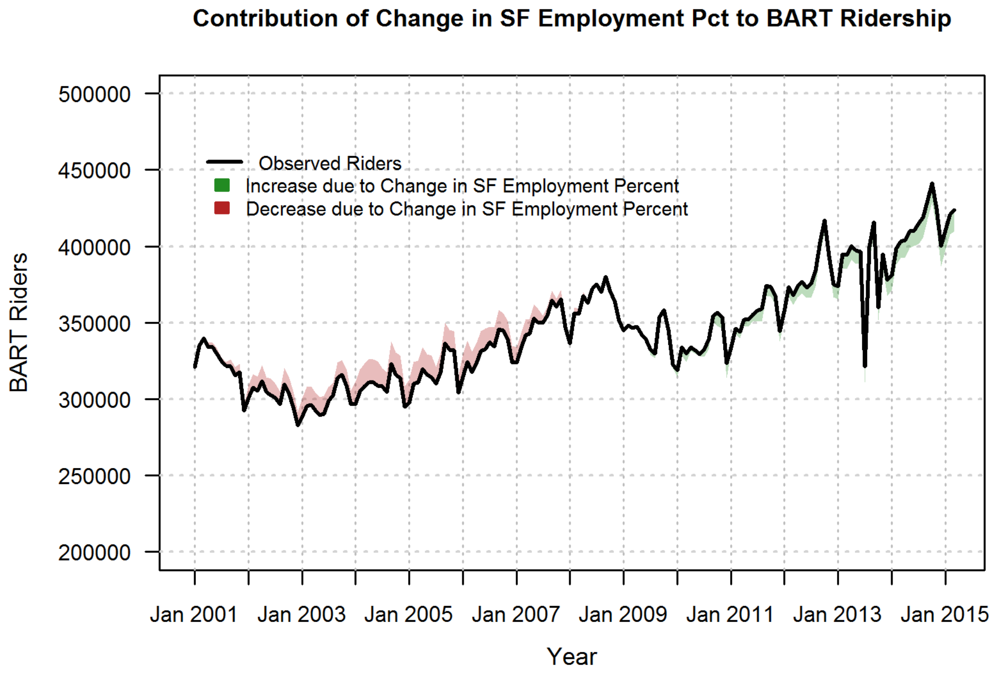


Figure 5.28: Effect of changes in percent of employment in San Francisco on BART ridership, vs. Jan 2001

Figure 5.29 shows the effect of changes in employment concentration on BART ridership over the 2009 to 2013 period. Over this period, employment becomes more concentrated, contributing to a relative increase in ridership.

5.4.4.5 Fare

Figure 5.30 and Figure 5.31 show the change in ridership attributable to fare changes. Fares are relatively stable over both periods, so the changes are not large.

5.4.4.6 Car Fuel Cost

Figure 5.32 and Figure 5.33 show the change in ridership attributable to changes in car fuel cost. Fuel cost increases contribute to a small ridership increase.

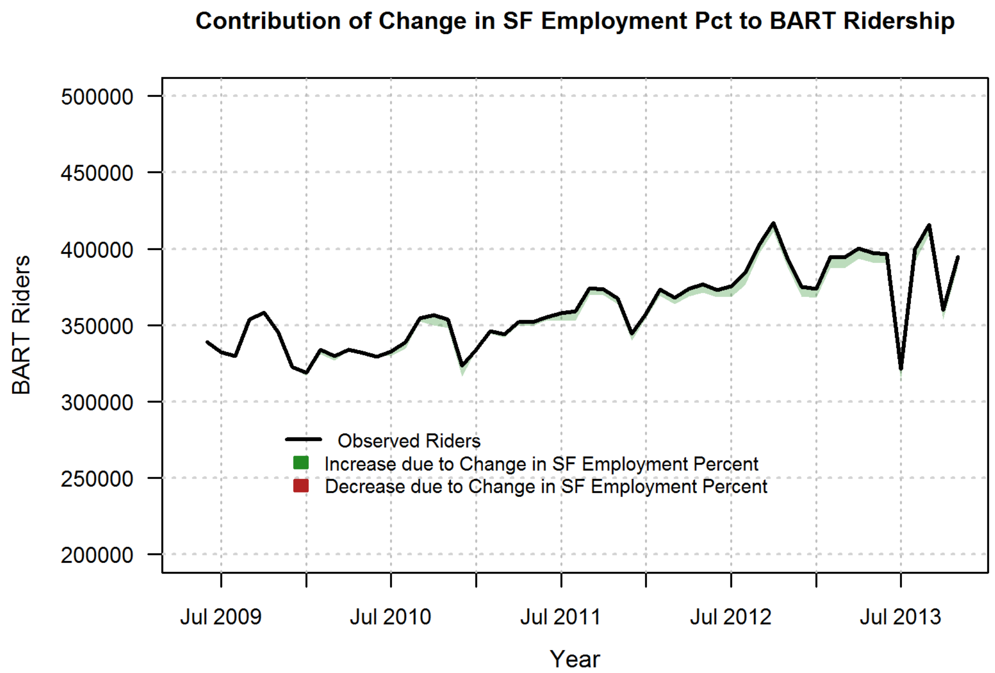


Figure 5.29: Effect of changes in percent of employment in San Francisco on BART ridership, vs. Sep 2009

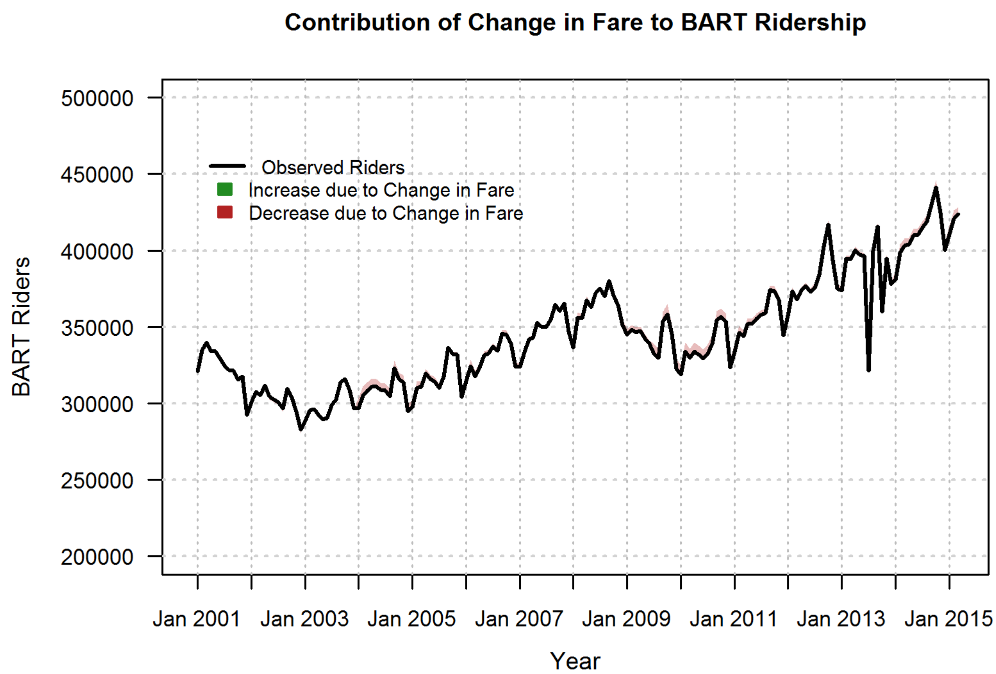


Figure 5.30: Effect of changes in fare on BART ridership, vs. Jan 2001

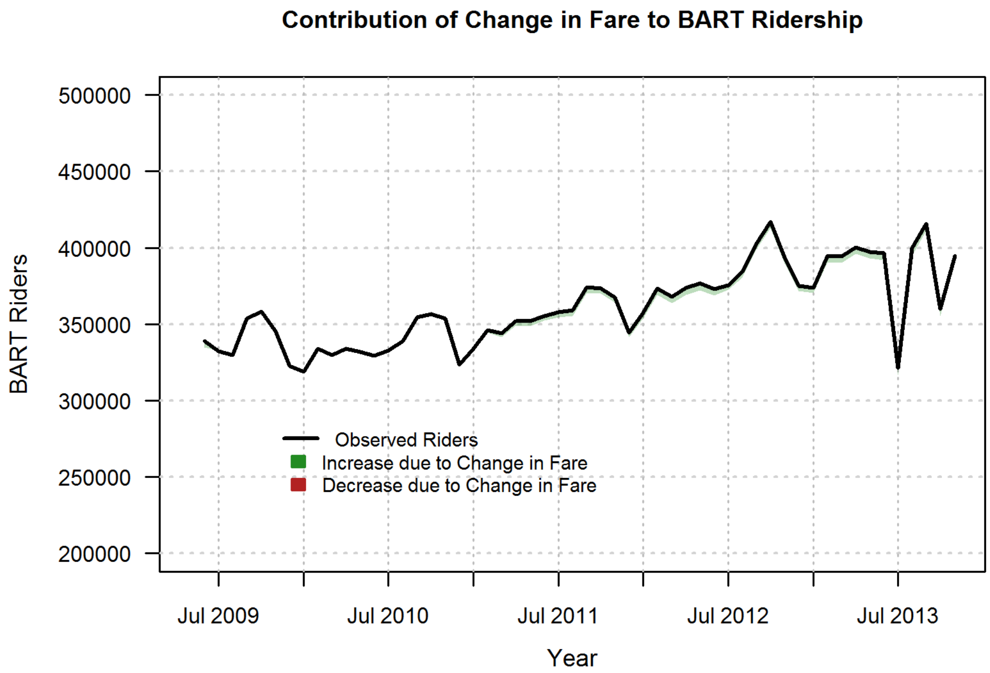


Figure 5.31: Effect of changes in fare on BART ridership, vs. Sep 2009

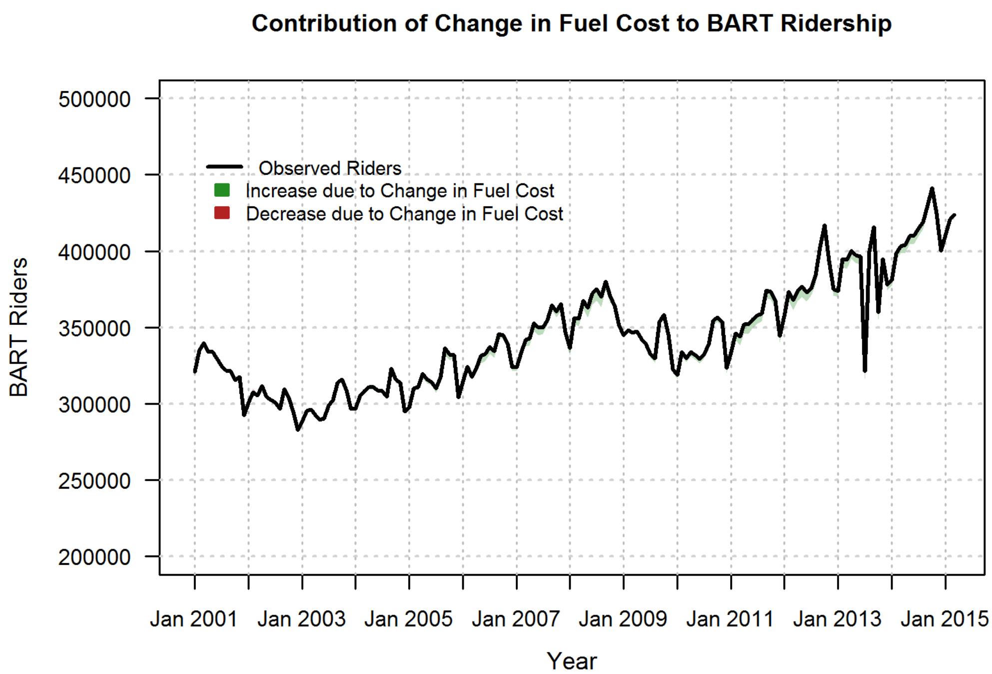


Figure 5.32: Effect of changes in car fuel cost on BART ridership, vs. Jan 2001

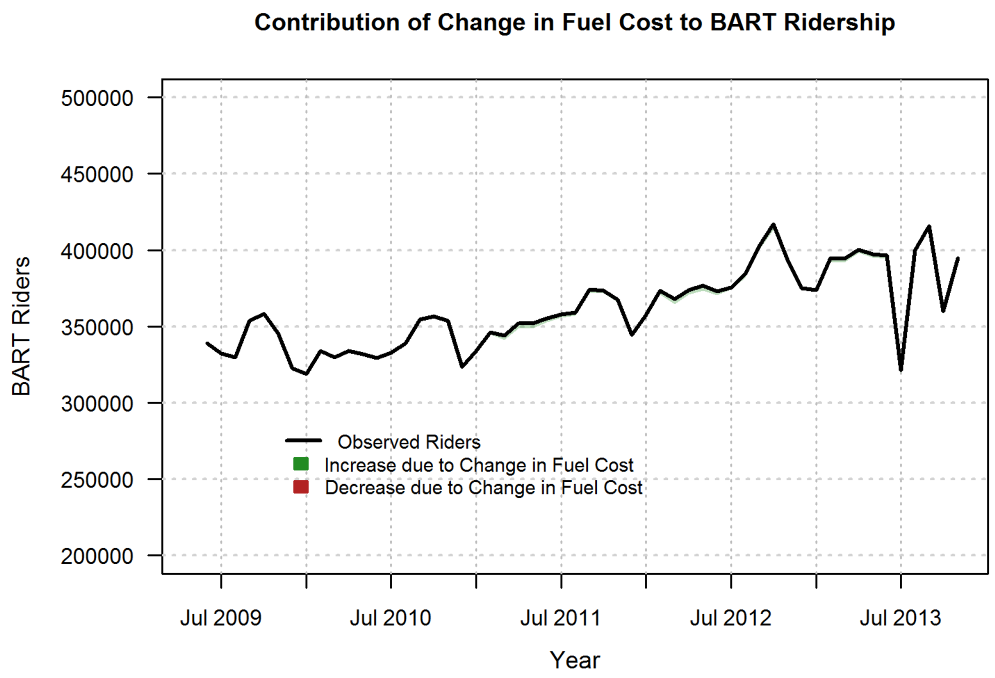


Figure 5.33: Effect of changes in car fuel cost on BART ridership, vs. Sep 2009

5.4.4.7 Strikes

Figure 5.34 and Figure 5.35 show the effect of strikes on BART ridership. The effect is limited to the two months in which there are strikes, and indicates that the time series would otherwise follow the ongoing trend.

5.4.4.8 Unexplained Trend

Figure 5.36 shows the change in BART ridership that cannot be explained by either the regression variables included in the model, or by random error.

There is a strong, and generally increasing unexplained trend over this period, that appears to be larger after 2009 than before. At the end of this period, BART has about 70,000 more riders than can be explained by the terms included in the model. This is a 20% higher than what it otherwise would have been.

While it may be tempting to think that the growth attributable to this trend will continue to increase, focusing on the 2009 to 2013 period, as shown in Figure 5.37 shows that the unexplained trend is not a major contributor to change over this period. Rather, the unexplained trend over this period mostly smooths the peaks

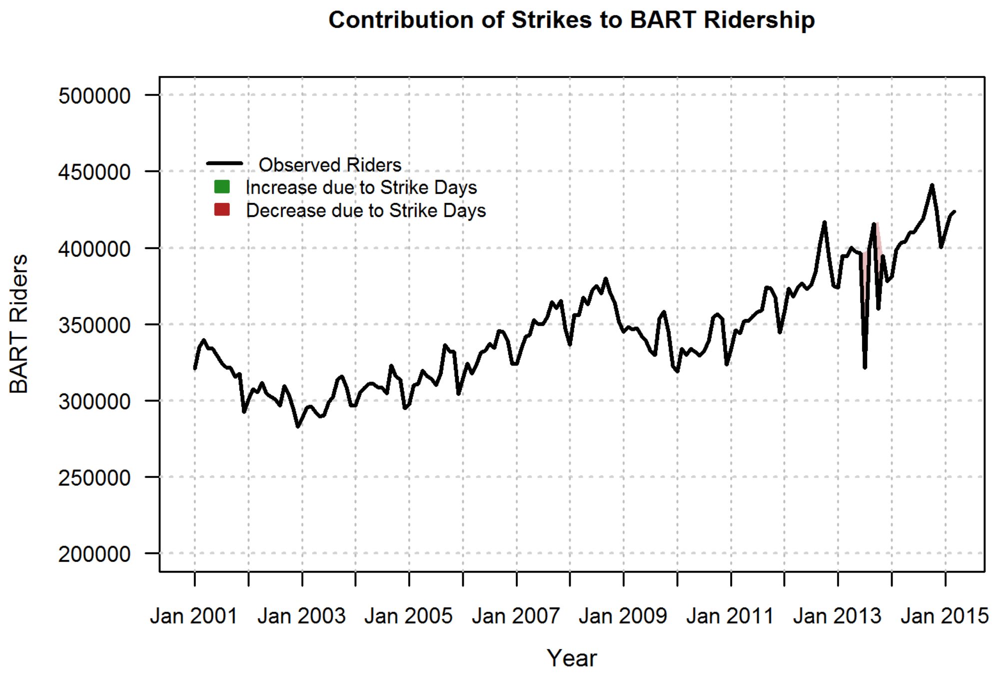


Figure 5.34: Effect of strikes on BART ridership, vs. Jan 2001

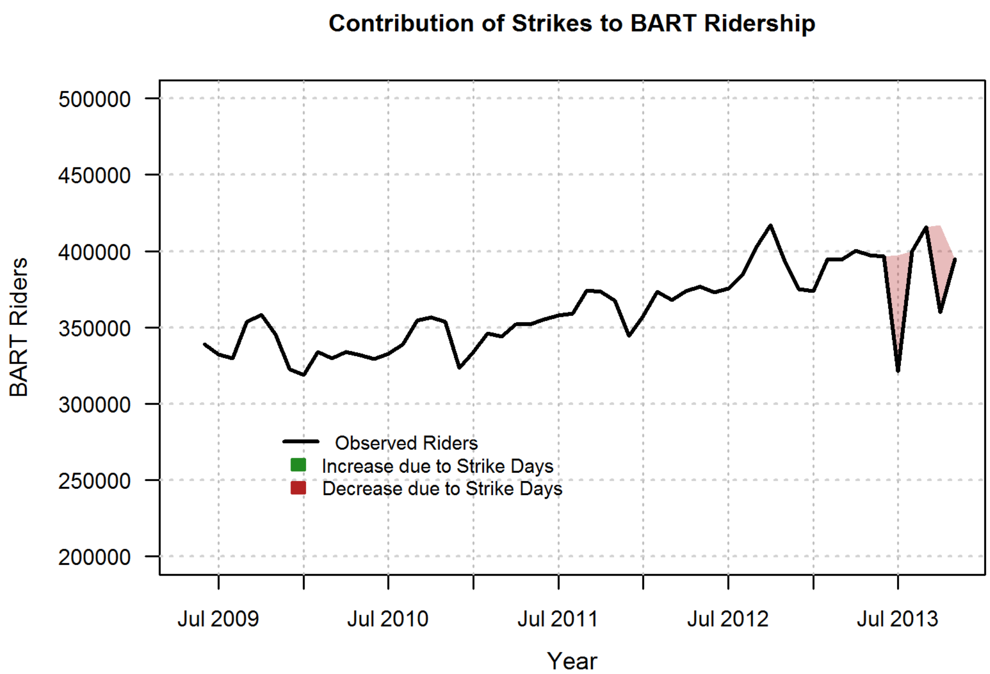


Figure 5.35: Effect of strikes on BART ridership, vs. Sep 2009

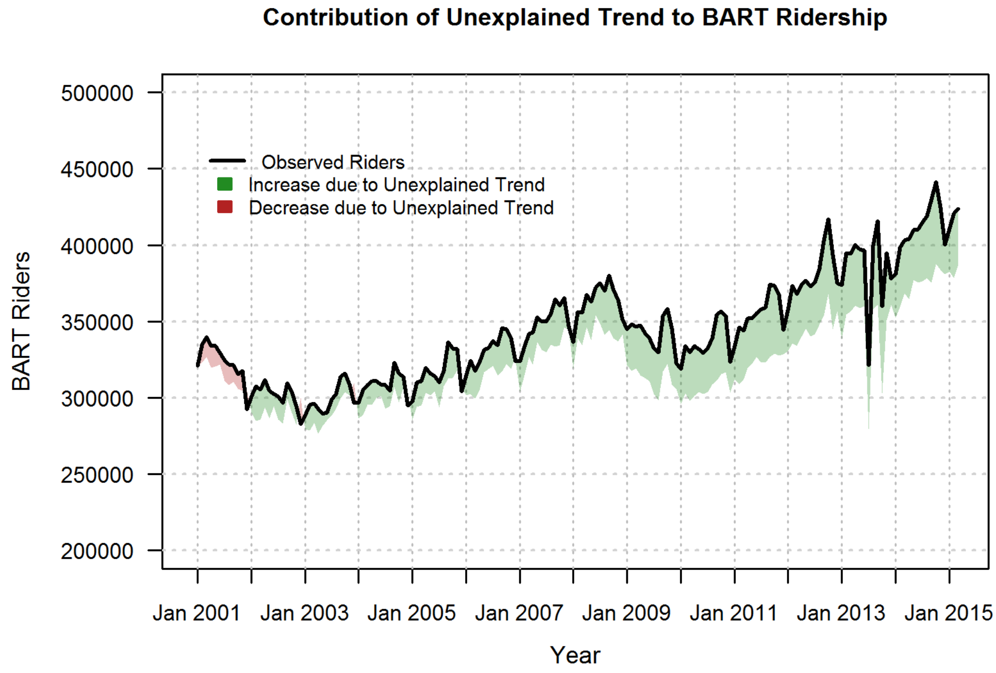


Figure 5.36: Effect of unexplained trend on BART ridership, vs. Jan 2001

and valleys associated with seasonality.

In mentally rectifying the two plots, this means that the width of the green shading in Figure 5.36 is relatively constant from September 2009 through 2013.

5.4.4.9 Residual Error

Figure 5.38 and Figure 5.39 show the effect of changes in the residual error on BART ridership. The first plot shows that the errors are minimal, while the second shows a general decrease. The latter is the change in e_t from September 2009, so it implies that there happens to be an especially large residual in that particular month.

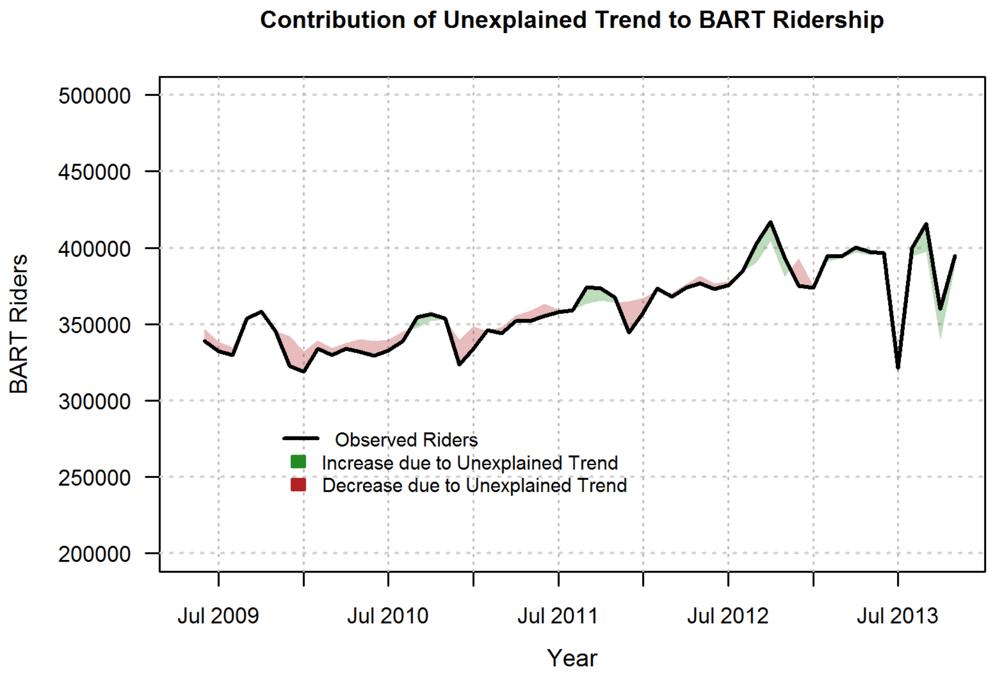


Figure 5.37: Effect of unexplained trend on BART ridership, vs. Sep 2009

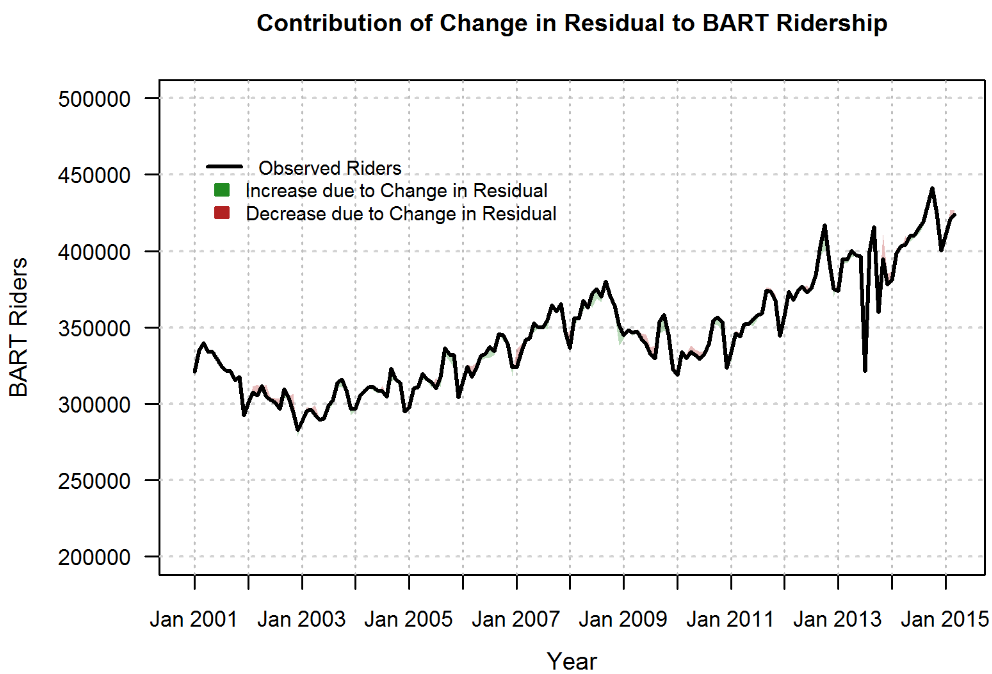


Figure 5.38: Effect of residuals on BART ridership, vs. Jan 2001

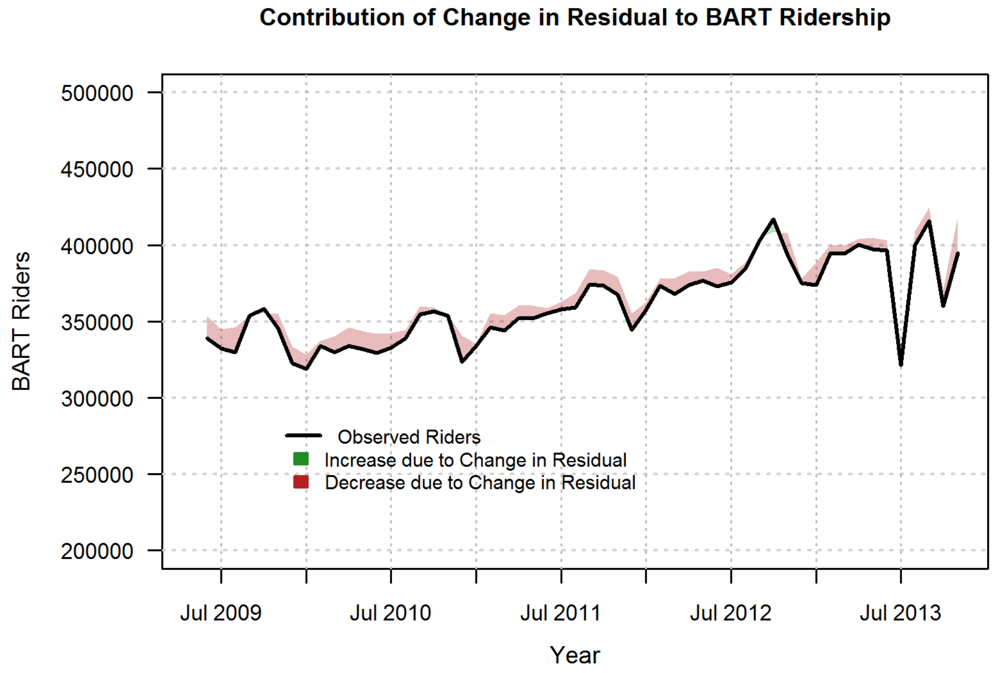


Figure 5.39: Effect of residuals on BART ridership, vs. Sep 2009

5.4.5 Comparison for 2009 to 2013 Period

The factors contributing to the change in MUNI and BART ridership between September 2009 and September 2013 are examined in further detail here. The goal of this analysis is to explain why BART ridership increases while MUNI ridership decreases, both during a period of economic growth. The September beginning and end points factor out the effect of seasonality, and covers a "typical" month when schools are in session and there are no strikes.

Table 5.6 shows the contribution of each factor in the model to the change in MUNI ridership, and Table 5.7 shows the contribution of each factor to the change in BART ridership. Over this period, MUNI ridership decreases by 35,000 riders, or 6.5%, while BART ridership increases by 62,000 riders, or 17.6%.

Service cuts reduce MUNI ridership by 4.6%. However, that change is partially offset, because rail service is also cut, shifting some riders back to bus. Lower runspeeds, from almost 11 mph in September 2009 to 10.5 mph in September 2013, contribute another 4.3% ridership reduction. Together, these three service changes

Table 5.6: Contributions to change in MUNI ridership: Sep 2009 to Sep 2013

Description	Lag	Coef	Value Sep-09	Value Sep-13	Ridership Change Absolute	Ridership Change Percent
Weekday service miles, 1000s		7,971	57.75	54.62	-24,995	-4.6%
Weekday service miles on MUNI rail, 1000s		-2,777	16.71	12.43	11,892	2.2%
Average bus runspeed		49,853	10.97	10.50	-23,431	-4.3%
Employment in San Francisco		0.876	544,587	615,119	61,786	11.3%
Unexplained trend			-890,620	-967,172	-76,552	-14.0%
Residual			-87	15,875	15,962	2.9%
Total Ridership			547,166	511,830	-35,337	-6.5%

Table 5.7: Contributions to change in BART ridership: Sep 2009 to Sep 2013

Description	Lag	Coef	Value Sep-09	Value Sep-13	Ridership Change Absolute	Ridership Change Percent
Weekday service miles, 1000s	D(0,12)	2712	28.56	27.37	-3,224	-0.9%
Number of Stations	D(0,12)	5472	43.00	44.00	5,472	1.5%
Employment in 4-county area		0.2027	1,812,112	1,993,007	36,667	10.4%
Percent of 4-county employment in SF		8099	30.05	30.86	6,570	1.9%
Cash fare (2010 $)		-20795	3.57	3.34	4,803	1.4%
Average car fuel cost (2010 $/mile)		86312	0.14	0.16	1,841	0.5%
Days with a BART strike		-19010	0.00	0.00	0	0.0%
Unexplained trend			-514,832	-496,144	18,688	5.3%
Residual			7,477	-1,217	-8,694	-2.5%
Total Ridership			353,681	415,805	62,124	17.6%

add up to a 36,500 ridership reduction, or 6.7% of the total. This is slightly more than the total change, indicating that if it were not for the service cuts, MUNI ridership would be nearly identical to its starting value.

In contrast, cuts to BART service over this period correspond to a slight decrease in ridership, but this is offset by one additional station that opens. The net effect of BART service and infrastructure changes is less than 1%.

Over this period, the employment in San Francisco grows from 545,000 to 615,000, or about 13%. The model predicts that this is corresponds to a 11.3% increase in MUNI ridership. Over this same period, the employment in the 4-county area grows by 10%, resulting in a 10.4% increase in BART ridership. Employment becomes more concentrated in San Francisco County over this period, increasing from 30.05% of the 4-county total to 30.86% of the 4-county total. The employment concentration contributes an additional 1.9% to BART ridership.

During this period, the average BART fare goes down from \$3.57 to \$3.34. The value is inflation adjusted to 2010 dollars, and the decrease reflects that inflation adjustment. BART fares were increased in July 2009, shortly before this analysis period. The model predicts that this change relates to a 1.4% increase in BART ridership. MUNI fares also increased in July 2009, and were stable over this period except for the inflation adjustment. A cost coefficient could not be estimated for the MUNI model, but its effect is be expected to be small over this period.

The average cost of fuel increases from \$0.14 per mile to \$0.16 cents per mile. This reflects an increase in the price of gasoline, that is partially offset by an increase in the average fuel economy of vehicles. This change results in a small increase in BART riders, of about 0.5%.

There are no strikes in either month, so the BART strikes have no effect on the ridership change.

The biggest drag on MUNI ridership is the unexplained trend, which serves to reduce MUNI ridership by 77,000 or 14%. In contrast to MUNI, The BART model has an unexplained trend over this period that results in BART ridership increasing by 18,000, or about 5.3%. The residual error contributes +2.2% to MUNI ridership and -2.5% to BART ridership.

It is not clear what is causing the large downward trend in MUNI ridership, other than that it is not something fully reflected in the model. It is interesting to note that the trend more than offsets the ridership growth due to the increase in employment. Tests of the MUNI model with an unconstrained (and much higher) employment coefficient showed that the unexplained trend in that model was more negative, still offsetting the ridership growth associated with employment growth, so the two appear to be related in opposite directions. One possible explanation is that MUNI ridership is simply insensitive to employment growth. Another is that MUNI ridership is sensitive to employment growth, but there is some other factor influencing ridership at the same time. These possibilities are considered, as the hypotheses are re-visited in the next section.

5.5 Hypotheses Revisited

At the start of this chapter, several hypotheses were proposed that might explain why BART ridership grows while MUNI ridership sputters, given strong economic conditions for both. Those hypotheses are revisited here to examine whether the evidence of the time series models is consistent or inconsistent with those explanations.

Table 5.8 summarises and compares the major factors contributing to changes in ridership for these two systems, grouped into slightly more aggregate categories than the tables above. This table provides the basic explanation for why BART ridership increases while MUNI ridership does not, and provides the basis for discussion of the hypotheses.

Table 5.8: Comparison of factors contributing to changes in MUNI vs. BART riders: Sep 2009 to Sep 2013

	Change in MUNI Riders		Change in BART Riders	
Associated with change in:	Absolute	Percent	Absolute	Percent
Service and fare changes	-48,425	-8.9%	7,051	2.0%
Employment and employment concentration	61,786	11.3%	43,238	12.2%
Change in MUNI rail service	11,892	2.2%	0	0.0%
Increased cost of auto travel	0	0.0%	1,841	0.5%
Unexplained trends and random error	-60,589	-11.1%	9,994	2.8%
Total Change	-35,337	-6.5%	62,124	17.6%

5.5.1 Different Level-of-Service Trends

The first possible explanation was that the divergent ridership trends could be explained by differences in the level-of-service provided by the operators. The analysis indicates that this explains part, but not all, of the divergence. Service cuts (changes to service miles and runspeed) between 2009 and 2013 result in a 8.9% loss of MUNI ridership. On BART, there is a slight reduction in service miles, which is offset by the opening of one new station and a decline in the real value of BART fares. Together, these factors result in a 2.0% positive contribution to BART ridership.

5.5.2 Suburban Growth Exceeds Urban Growth

Our next hypothesis was that suburban growth exceeded urban growth, contributing to a faster increase in BART ridership because BART serves more suburban areas.

Examining the growth of employment, population and workers showed that employment became more concentrated in San Francisco over the analysis period, and that there was little change in the concentration of population and workers in San Francisco versus the outlying counties.

The model analysis showed that the strong economy, and associated growth in employment and the concentration of employment in San Francisco contribute to ridership gains on both systems. The gains are slightly greater for BART (12.1%) than for MUNI (11.3%).

There is some uncertainty in the estimate for MUNI, because the model considered here reflects a constrained employment coefficient. Leaving the coefficient unconstrained results in an estimate 2.5 times as high, which would suggest a larger growth in ridership associated employment growth. However, in both cases, this is offset by an unexplained downward ridership trend. This could indicate either that MUNI ridership is less sensitive to employment growth than indicated here, or that there is some other factor causing a relative ridership decline. There are both theoretical reasons to believe that transit ridership should be tied to employment, and enough empirical evidence from other studies [174, 129, 173, 180] to suggest that it is, that we are inclined to believe the latter, even if the magnitude is uncertain.

From this analysis, we do not find evidence that suburban growth exceeding urban growth is a major factor driving the divergence.

5.5.3 MUNI Trips Shift to Rail

Another possible explanation was that there was a shift of MUNI bus trips to the MUNI rail system. The analysis was inconsistent with this hypothesis, and appears to indicate an effect in the opposite direction. Rail service was cut during this period, contributing to a relative increase in MUNI bus ridership.

5.5.4 Increased Cost of Car Travel

One reason BART ridership would increase faster than MUNI ridership would be if the cost of auto travel increased. Because BART trips tend to be longer, they are likely to be more sensitive to changes in auto travel costs.

The analysis showed that the increases to the average car fuel cost does increase BART ridership, and no equivalent factor could be estimated for MUNI. However, the effect is small over the 2009 to 2013 period, contributing less than 1% to the increase in BART ridership.

With the data available, we were not able to find a significant relationship between aggregate car congestion measures and BART or MUNI ridership.

5.5.5 Transit Riders are Priced out of San Francisco

This research considered the possibility that as San Francisco becomes a more expensive city, transit riders are being priced out of the city, and into other counties.

This would be logical, given that net growth in households in San Francisco in recent years has been exclusively among households earning $100,000 or more per year, and higher income travellers tend to be less likely to use transit. However, it is also observed that transit commute mode shares among high-wage workers have increased, and are now about as high as among lower wage workers.

To test whether there is evidence that income changes were contributing to the stagnation of MUNI ridership, a number of income related variables were tested in model estimation, both for the MUNI and BART models. These included households segmented by income, workers and employees segmented by wages, and the median household income. These tests produced either insignificant or illogical model parameters.

The end result is that this analysis does not provide evidence to support the hypothesis that transit riders are being priced out of San Francisco.

5.5.6 The "Uber Effect"

Another hypothesis is that there is an "Uber Effect", whereby new or alternative modes become attractive enough to draw substantially from MUNI ridership. The journey to work data show a marked upswing in the share of workers commuting by walk, and commuting by bike, taxi and other. To test this hypothesis, descriptive variables were included in the models for the number and share of workers commuting by walk, bike, taxi and other modes. The estimated coefficients were negative, as expected, but insignificant. It is logical that it would not affect BART ridership, because BART tends to attract longer regional trips for which these modes would be less competitive, but a relationship with MUNI may be reasonable to expect.

To examining the relationship further, Figure 5.40 shows the unexplained trend from the MUNI model in comparison to the trend in walk, bike, taxi and other commute mode share. The unexplained downward ridership trend corresponds to an upward trend in walk, bike, taxi and other commutes, but it is a rough relationship. The analysis is hampered by the fact that measurements of the commute mode share are aggregate in time because they are based on an annual survey, and that they are an imperfect proxy for total travel by these modes. In fact, the Shared Use Mobility Center (SUMC) suggests that trips by shared mobility services tend not to be work trips, and that social trips which can substitute for drunk driving are a particularly important market [221]. SUMC argues that for these reasons, shared mobility modes tend to be more complementary than competitive with transit systems.

While it may be possible that there is an “Uber Effect” that contributes to the unexplained trend in MUNI ridership, the data and relationships do not provide sufficient evidence to make a case that it is present.

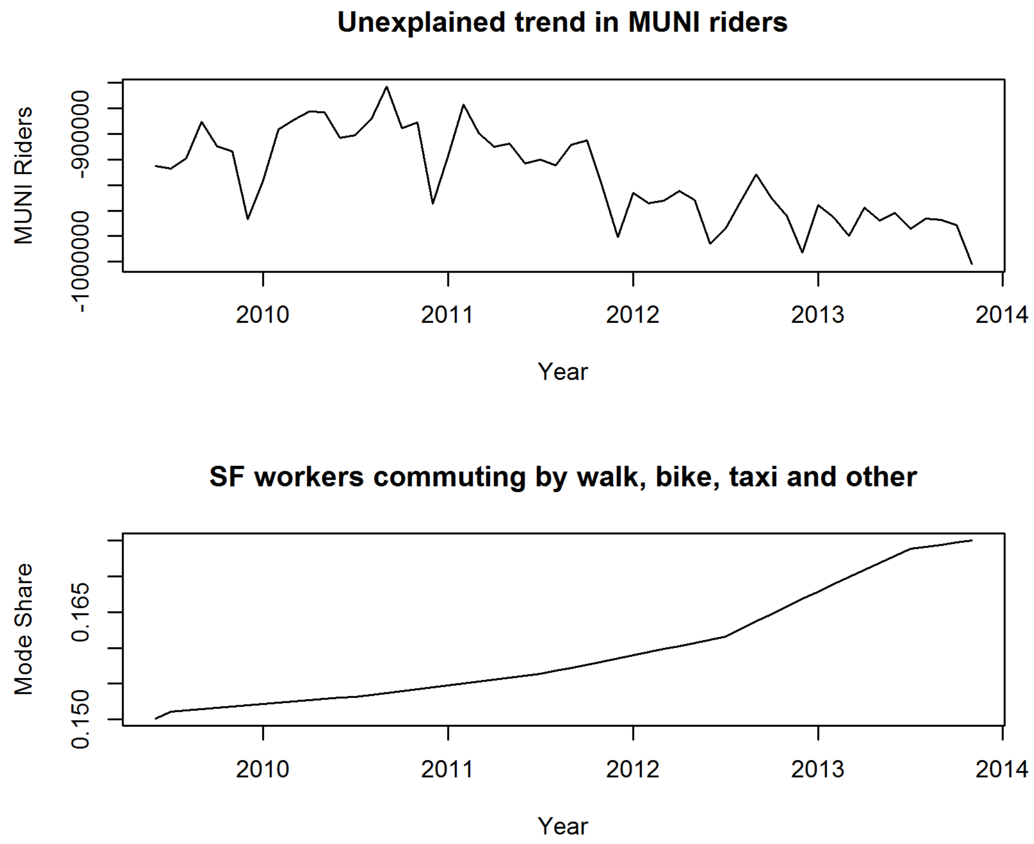


Figure 5.40: Unexplained trend in MUNI ridership vs. walk, bike, taxi and other commuters

5.6 The Effect of an Ageing Population

Having re-considered our original hypotheses given the evidence provided by the application of the time series models, we do not find the explanations to be fully satisfactory in terms of explaining the ridership divergence between BART and MUNI. Specifically, the models do a reasonably good job of explaining the trends in BART ridership, but there remains an large downward trend in MUNI ridership that the models cannot explain. Therefore, we consider one more possible explanation: that the changing composition of the population by age group could negatively affect MUNI ridership more than BART ridership. This explanation builds from the trends found in the UK where older women now drive more than in past cohorts while younger men drive less than in past cohorts [183, 184]. Could it be that a similar trend is affecting bus ridership?

This section considers that possibility in two parts. First it describes a series of additional models estimated to test age effects, and next it considers the ridership implications of these revised models.

5.6.1 Model Estimation

A series of additional RegARIMA models were estimated to consider the effects of age on BART and MUNI ridership. These models do not explicitly consider the cohort effects, but rather the aggregate effect of changes in the age composition of the population. The starting point is the preferred RegARIMA model of BART ridership (Table 4.16), and the preferred RegARIMA model of MUNI ridership with a constrained employment term (Table 4.15). For each, five new variables were considered, measured either for 4-county area for BART, or for San Francisco County for MUNI. The variables tested include: median age, percent of the population age 20 to 29, percent of the population that is male and age 20 to 29, percent of the population age 65+ and percent of the population that is female and age 65+. These variables were selected to consider the possibility that there is a different travel behaviour by young adults, specifically by young males, by older adults, or specifically by older women. Combinations of these variables were also tested, such as both the young adults and older adults terms. In all cases, the data are taken from the annual American Community Survey (ACS), and interpolated to monthly values.

For the BART models, only one of the terms estimated gives a significant coefficient: the percent of the population age 65+. It has a positive coefficient of about 46,000, indicating that for each 1% increase (such as from 14% to 15%) in the share of the population age 65+, there is an increase of 46,000 riders. The trade-off to including this new variable is that the coefficient on service miles becomes small and insignificant. There appears to be co-linearity between these two terms, making it difficult to separate the effects. We would find it strange to have a model that is not sensitive to the service provided, so prefer the previous BART model.

For MUNI, the best model is one that includes the percent of the population age 65+. The percent of the population age 20 to 29 was also significant on its own, but when the two terms are included together, they both become insignificant.

Table 5.9 shows the results of this estimation, with the addition of a term on the percent of the population aged 65+, and with the employment term still fixed. A direct comparison against the previous estimation results is shown later in this section. The interpretation of the age term is such that a 1% increase in the percent of the San Francisco population age 65+ (such as from 14% to 15%) is associated with a decrease of 155,000 MUNI bus riders. This is a large number, so it should be noted that the data vary within a narrow range over the estimation period, from a low of 13.7% to a high of 14.3%. This result should be used with a dose of caution, especially if the values change substantially in future applications.

Table 5.9: RegARIMA model of MUNI boardings with constrained employment term and age

Model Characteristics				
Dependent variable	MUNI boardings			
Type	$ARIMA(0,1,1)(0,1,0)_{12}$			
Date range	Jun 2009 to Nov 2013			
Predictive Variables				
Description	Lag	Coefficient	S.E.	T-Stat
Moving average coefficient	1	-0.4516	0.1994	-2.26
Weekday service miles, 1000s		11432	3203	3.57
Weekday service miles on MUNI rail, 1000s		-3579	2248	-1.59
Average bus runspeed		37859	26678	1.42
Employment in San Francisco		0.876	fixed	fixed
Percent of population age 65+		-155305	72339	-2.15
Model Statistics				
Log likelihood		-451.62		
AIC		915.24		
AICc		917.71		
RMSE		14,667		
Percent RMSE		2.94%		
Box-Pierce test p-value		0.622		

The model residuals remain stationary, as confirmed by Figure 5.41, and the Box-Pierce test shown in Table 5.9. In addition, the Ljung-Box gives a p-value of 0.49. Figure 5.42 shows the modelled and observed time series plots, and Figure 5.43 shows the resulting scatterplots.

Table 5.10 shows a comparison of the model estimation results with and without the age term. The employment coefficient is fixed, and remains unchanged. The

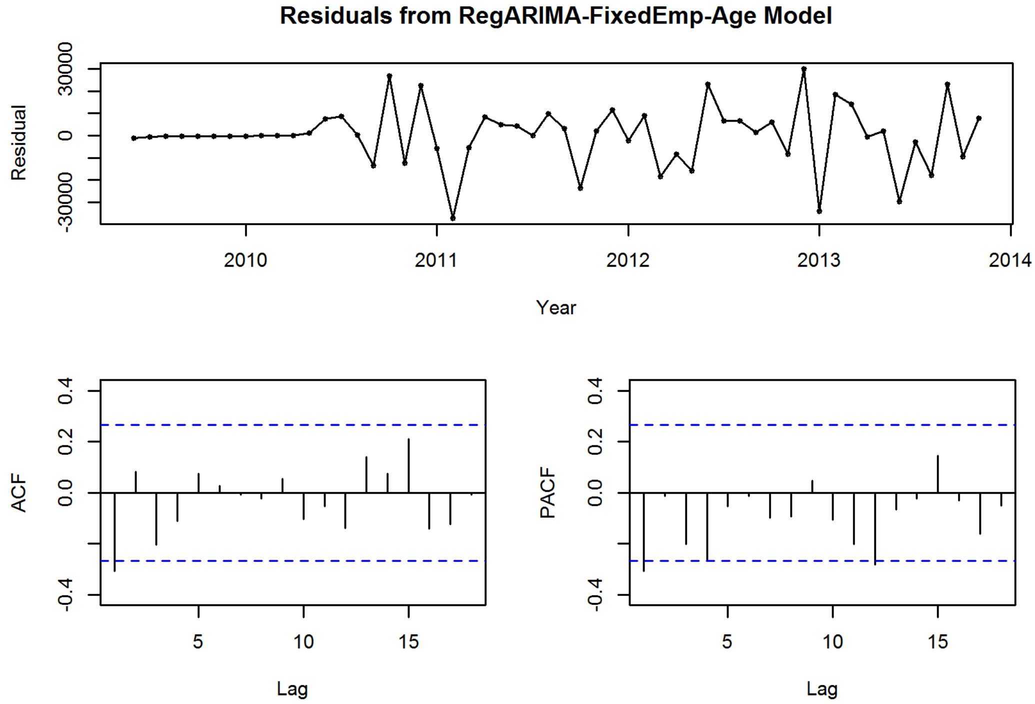


Figure 5.41: Residual autocorrelation from MUNI RegARIMA model with constrained employment term and age

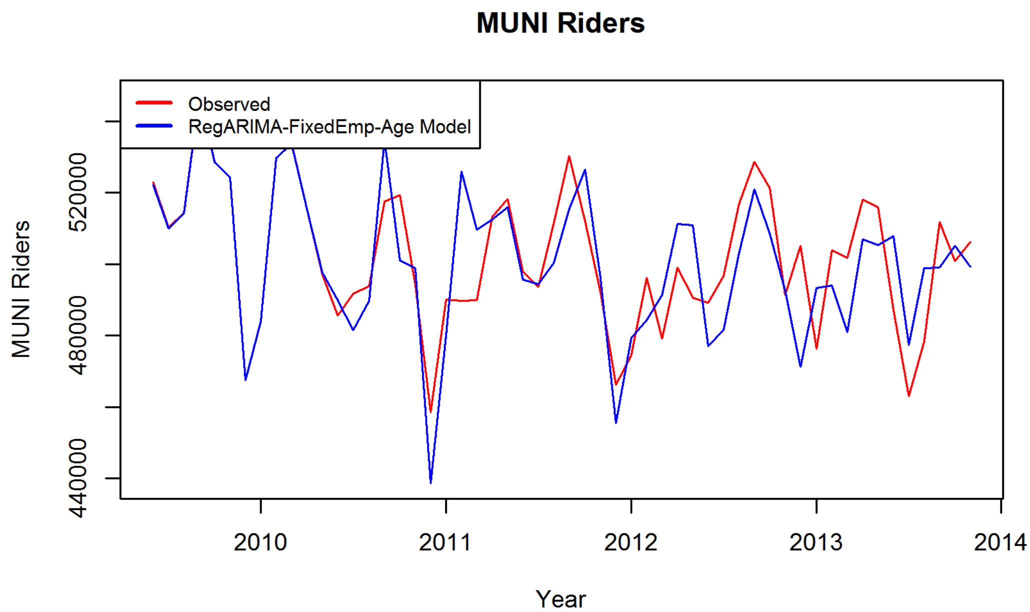


Figure 5.42: MUNI boardings, observed vs. RegARIMA model with constrained employment term and age

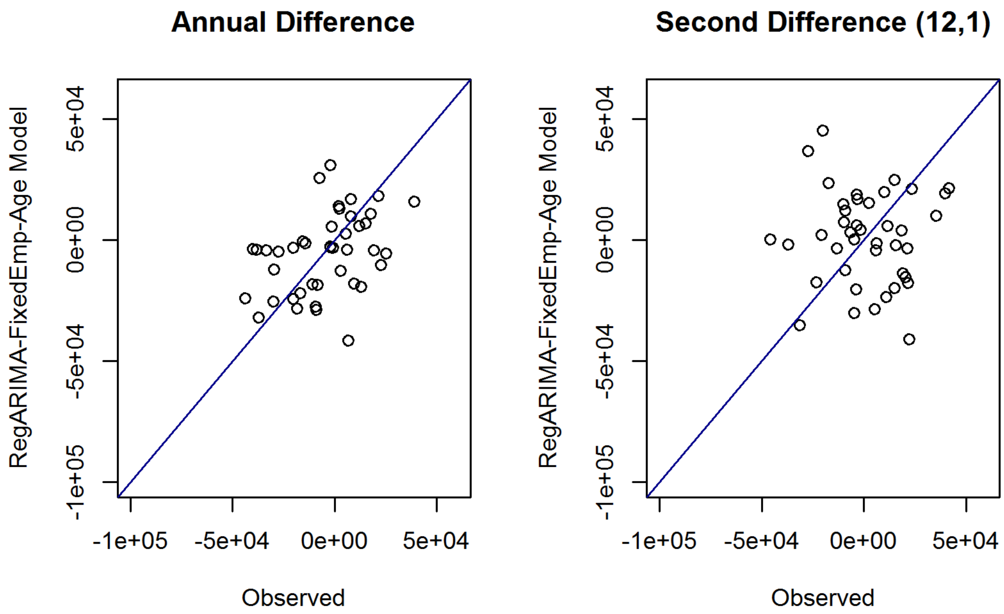


Figure 5.43: Change in MUNI boardings, observed vs. RegARIMA model with constrained employment term and age

moving average coefficient, weekday service miles and MUNI rail service miles increase in magnitude and significance, while the average bus runspeed decreases in magnitude and significance. In spite of this lower significance, the runspeed term is retained in the model for consistency with the previous models and for its policy sensitivity. The goodness of fit measures improve for the model with the age term. Overall, the model appears to be improved with the addition, although the large magnitude of the age coefficient should lead us to be cautious when applying the model if there are large changes in the percent of the population aged 65+.

5.6.2 Ridership Changes

In this section, the updated model is applied to quantify the contributors to the change in MUNI ridership between September 2009 and September 2013, as shown in Table 5.11. These values are then compared to the calculations with the base model, shown in Table 5.12.

The major difference between these two sets of calculations is that the large unexplained trend in the base model is partially explained by the effect of an ageing

Table 5.10: Comparison of MUNI RegARIMA Models with and without age term

Model Characteristics					
Dependent variable	MUNI boardings				
Date range	Jun 2009 to Nov 2013				
Predictive Variables					
		Base Model		With Age Term	
Type		$ARIMA(0,1,1)(0,1,0)_{12}$		$ARIMA(0,1,1)(0,1,0)_{12}$	
Description	Lag	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1	-0.3092	-1.67	-0.4516	-2.26
Weekday service miles, 1000s		7971	2.57	11432	3.57
Weekday service miles on MUNI rail, 1000s		-2777	-1.12	-3579	-1.59
Average bus runspeed		49853	1.94	37859	1.42
Employment in San Francisco		0.876	fixed	0.876	fixed
Percent of population age 65+				-155305	-2.15
Model Statistics					
Log likelihood		-453.56		-451.62	
AIC		917.12		915.24	
AICc		918.83		917.71	
RMSE		15,401		14,667	
Percent RMSE		3.09%		2.94%	
Box-Pierce test p-value		0.873		0.622	

population in the revised model. Specifically, the base model has a -14% unexplained trend, compared to the revised model with a -4.8% unexplained trend plus a -8.2% trend explained by the age term, for a combined effect of -13%. Thus, it appears that this age term can explain about 60% of the previously unexplained decline in MUNI ridership.

This appears to be a valuable improvement to the model and to the analysis. However, there is reason to remain sceptical, because the percent of the population aged 65+ only changes from 14.0% to 14.3% over this three year period. Can such a small change in age composition really affect transit ridership that much? In examining this question, we consider two additional items. First, the ACS, which is the source of the age share data, provides margins of error on their estimated values. The margin of error on the percent of the population aged 65+ in San Francisco is 0.1%, so the change of 0.3% is beyond the margin of error. Second, we consider the change in trips relative to the actual number of people in the age 65+ group. Between September 2009 and September 2013, the population aged 65+ increases from 113,604 to 119,698, for an increase of 6,094 persons in that age group. According to Table 5.11, the age variable is associated with a decrease of 44,945 MUNI boardings. This implies that for each additional person aged 65+, there are 7.4 fewer MUNI boardings. Given that there are only 0.6 MUNI average weekday boardings for each person living in San Francisco in 2013, the rate of 7.4 fewer boardings per person aged 65+ appears to be unreasonably high.

While is difficult to be certain of the quantity, it does appear that population composition effects, such as an ageing population, may be an important factor to consider when predicting transit demand.

Table 5.11: Contributions to change in MUNI ridership, considering age effect: Sep 2009 to Sep 2013

			Value		Ridership Change	
Description	Lag	Coef	Sep-09	Sep-13	Absolute	Percent
Weekday service miles, 1000s		11,432	57.75	54.62	-35,848	-6.6%
Weekday service miles on MUNI rail, 1000s		-3,579	16.71	12.43	15,326	2.8%
Average bus runspeed		37,859	10.97	10.50	-17,794	-3.3%
Employment in San Francisco		0.876	544,587	615,119	61,786	11.3%
Percent of population age 65+		-155,305	14.0	14.3	-44,945	-8.2%
Unexplained trend			1,222,445	1,195,976	-26,469	-4.8%
Residual			172	12,779	12,607	2.3%
Total Ridership			547,166	511,829	-35,337	-6.5%

Table 5.12: Comparison of contributions to change in MUNI ridership, with and without age effect: Sep 2009 to Sep 2013

	Base Model		With Age Term	
Description	Absolute	Percent	Absolute	Percent
Weekday service miles, 1000s	-24,995	-4.6%	-35,848	-6.6%
Weekday service miles on MUNI rail, 1000s	11,892	2.2%	15,326	2.8%
Average bus runspeed	-23,431	-4.3%	-17,794	-3.3%
Employment in San Francisco	61,786	11.3%	61,786	11.3%
Percent of population age 65+			-44,945	-8.2%
Unexplained trend	-76,552	-14.0%	-26,469	-4.8%
Residual	15,962	2.9%	12,607	2.3%
Total Ridership	-35,337	-6.5%	-35,337	-6.5%

5.7 Conclusions

This research has considered recent divergent trends in ridership on two San Francisco Bay Area transit systems. Between 2009 and 2013, BART ridership has increased 17%, while MUNI bus ridership has decreased 6.5%. The research focuses specifically on the 2009 to 2013 period where the most detailed data are available, but these changes are part of a longer trend, where BART ridership has increased 50% between 1998 and 2013, while MUNI ridership is nearly identical to its 1998 value.

5.7.1 What is Driving the Ridership Divergence?

The preferred RegARIMA models were applied to determine the relative contribution of each descriptive variable to the change in ridership that occurred on each system between 2009 and 2013. This provided evidence in favour or opposition to six hypotheses that were proposed to explain the different trajectories of BART and MUNI ridership.

The research found that different level-of-service changes and a shifting jobs-housing balance were important contributing factors to the difference. Service and fare changes contributed to 9% less MUNI ridership and 2% more BART ridership.

With respect to land use characteristics, we had hypothesised that suburban growth exceeding urban growth would contribute to relatively higher growth in BART ridership. When measured at a county level, the households, population and workers do not become less centralised over this period, while the employment actually becomes more concentrated in San Francisco. Together, employment and its concentration in San Francisco drives a 11% increase in MUNI riders and a 12% increase in BART riders, while no additional effect could be found for households, workers and population. This data does not support the hypothesis of decentralised growth.

The research did not find evidence in support of the next three hypotheses. We hypothesised that MUNI bus trips had shifted to rail over this period, but found evidence to suggest that cuts in rail had instead increased bus ridership. Increases in car fuel cost do contribute to an increase in BART ridership, whereas no effect could be estimated for MUNI. The difference is small, though, contributing less than 1% additional BART ridership. Evidence was not found to support the hypothesis that MUNI riders are being priced out of the city as San Francisco becomes more expensive.

Similarly, the research did not find significant evidence in support of an "Uber effect", whereby a shift towards non-motorised and shared mobility modes causes a drag on MUNI demand. After controlling for all the factors included in the model, there is an unexplained downward trend in MUNI ridership, moving in the opposite direction from the upward trend observed in commutes made by walk, bike, taxi or other modes, but the relationship is weak.

The analysis shows that the biggest factor contributing to the divergent ridership is an unexplained trend. In combination with the residual error, this trend contributes to an 11% reduction in MUNI ridership versus a 3% increase in BART ridership over the analysis period.

Going back to the author's original motivation for this work, this analysis shows that most of the growth in BART ridership can be attributed to factors that should already be included in the regional travel demand model. It appears that the source of the 2013 under-prediction of BART ridership observed during the seismic retrofit project was most likely due to employment inputs in the model that did not reflect recent growth, rather than an underlying change in the behaviour of young people.

The interpretation of the unexplained trend is simply that something is causing this change beyond what is already accounted for in the model. The literature review (Section 5.2) provides some ideas from the types of explanations others have offered in understanding observed travel trends. It may be that the change relates to the introduction of a new mode (Uber and other shared mobility services) that cannot be picked up at the resolution of the data available. It could be a cultural shift in favour of active travel (walk and bike). It could be a demographic shift, or a shift in the spatial distribution of land use below the county level.

One possible explanation that was explored in further detail was the possible effect of the changing age composition of the population. Additional model estimations and applications showed that an increase in the percent of the population age 65+ in San Francisco can explain about 60% of the previously unexplained decline in MUNI ridership, with no similar effect on BART. While the small change in age distribution leads to some uncertainty in this estimate, it appears that an ageing population is an important factor to consider when predicting transit demand.

It is likely that there is some combination of factors causing the remaining downward trend in MUNI ridership. This trend is large enough that it has important implications to expected level of transit demand in the medium term, and warrants further research to better understand.

5.7.2 Future Research

There are a few directions that future research on the topic can go.

The first would be to extend this analysis to 2016 or beyond, to determine whether the trend persists, and whether a longer time series would allow significant estimates for a broader range of variables. The practical challenge here is bureaucratic rather than technical. There are delays in the release of certain data sources, such as the LEHD Origin-Destination Employment Statistics (LODES), and as staff contacts who have provided data in the past change jobs, it is necessary to invest in developing new relationships to gain access to other data.

Another direction is to attempt to replicate the findings in other, similar, cities. London and New York both come to mind as major cities with high transit mode shares, and technologically savvy populations that may be early adopters of shared mobility services, as well as markets for active travel.

Third, it may be beneficial to attempt a panel data approach, as suggested in the conclusions to Chapter 4.

5.7.3 Implications for Travel Forecasting

The large unexplained trends found in this research raise interesting questions about the stability of models over time, and whether the constants and other parameters can be expected to stay the same over time. One simple explanation is that the trend persists because some important variable is left out of the model, and if that variable were included the trend would be more stable. It may be that full scale travel models, estimated from disaggregate data, do a better in this regard because they are able to incorporate more variables. However, for models estimated from cross-sectional data, that data provides no indication as to the temporal stability of the models parameters.

Fox [233] examined this issue in the context of travel surveys collected across 20 years in Toronto and Sydney. He found that for mode and destination choice models, improving the model specification did tend to improve temporal transferability. Interestingly, models that accounted for tasted heterogeneity were found to have a better fit against cross sectional data, but did not necessarily do better in terms of temporal transferability.

A related means of addressing this issue may be to take advantage of multi-year travel surveys, such as the UK National Travel Survey (NTS). Mode choice models could be estimated with a mode-specific constant that varies by year. A constant that has a significant and monotonically changing trend across years may indicate that there is some underlying change not accounted for by the descriptive variables otherwise included in the model. Such a model would probably work best if the level-of-service data were also updated to be year specific, a process may become less cumbersome as data such as GTFS can be used to facilitate network development.

Broadly, this points to a situation where cross-sectional fit is not everything, and there is a value in incorporating data that varies over time into the travel forecasting process. Further work is needed to better capture and process recently available time-varying data, as well as to sort out how they can best be used to improve forecasts.

Chapter 6

Project-Level Applications

This thesis began, in Chapter 1, with a discussion of the value of ex-post evaluations of transport project demand. Such studies are useful to the field of travel demand forecasting, because they provide basis against which models can be compared for their ability to replicate past changes, and they build a body of reference cases against which to compare future forecasts [14].

Pratt [39] notes that most ex-post studies of transit change use a naive before-and-after approach that is limited because it does not account for the statistical noise in the data and can be highly dependent on the time periods chosen. The limiting factor in terms of applying even moderately sophisticated methods is often the cumbersome nature of assembling the required data, or in many cases, the lack of foresight to collect the right data in the first place [234].

Chapter 3 describes a data mashing tool developed in an effort to address this problem. It is focused on transit, and implemented for San Francisco as an example. The software tool accumulates continuously collected data from from several related sources, including the General Transit Feed Specification (GTFS) for the schedules, transit Automated Vehicle Location (AVL) and Automated Passenger Counter (APC) data, and transit smart card data. It combines them by expanding the less complete data sources to the more complete data. Data from other Bay Area transit systems are also included in the database, as are "drivers of demand" data. By drivers of demand, we mean those data that describe something about changes in the city that might drive changes in transit ridership. In addition, Chapter 3 noted the emerging trend towards performance based planning, and discusses how the data tool may be useful in facilitating performance based planning.

Chapter 4 took advantage of the outputs of this system, and estimated time series models of transit demand for the BART and MUNI systems in the San Francisco Bay Area. Chapter 5 went on use these models to understand the drivers of changes in transit demand.

This chapter goes further in linking these chapters through several example applications. It applies the models estimated in Chapter 4 for the purpose of ex-post evaluation, and it applies those same models to generate short term forecasts useful for establishing targets in the context of performance based planning. The goal is an approach that provides a quality answer for a low incremental cost of analysis.

The remainder of this chapter is structured as follows.

Section 6.1 starts by introducing three methods for estimating ex-post ridership changes. Then, three changes are evaluated with the time series based analysis: a set of systemwide service cuts to the San Francisco Municipal Railway (MUNI) bus system in 2010, a set of systemwide Bay Area Rapid Transit (BART) service cuts in 2009, and the 2003 extension of BART to San Francisco International Airport (SFO). For each change, the models are used to estimate the change in ridership attributable to the service changes. This result is compared to the result that would be obtained using two common methods of ex-post analysis: a naive before-and-after calculation, and the application of published elasticities.

Section 6.2 begins with a short introduction to performance based planning. Then, two examples are considered to demonstrate how short term forecasts from these models may be useful in establishing reference points for project evaluations. Both examples are pilot projects, where a target for a specific ridership change is considered prior to the pilot implementation. The examples show that such targets may be more or less difficult to achieve given pre-existing trends.

Finally, Section 6.3 summarises conclusions for both sets of analyses.

The contribution of this research is two-fold. First, it further highlights the limitations of evaluation methods commonly used in practice, and demonstrates how time series models can be used to mitigate those limitations. Second, it demonstrates how short term forecasts produced using the same set of time series models may be valuable in defining appropriate performance targets for planning projects.

6.1 Ex-Post Applications

In this section, three examples of transit system changes are considered, and three options are considered to measure the ridership change associated with each case.

The first is a before-and-after evaluation. In this approach, measurements are taken before a project is implemented, and after its implementation. The difference is the estimate of the change, sometimes with a qualitative discussion of potential confounding factors. The limitations of such an approach are well known, as discussed by Higgins and Johnson [235]. The basic problem is that it can remain unknown whether the change is attributable to the project, to some other external factor, or to random noise. In addition, Olsson et al. [57] demonstrated that the reference points selected for the before point-in-time and the after point-in-time can affect the conclusions in ex-post evaluations of major rail projects. In spite of these known limitations, such studies remain common, such as [29, 236, 237].

The second option is to apply elasticities reported in the literature. Such an approach is simple to apply and can provide a quick estimate of the expected change in demand associated with a change in service or price. Elasticities can be derived using time series analysis, such as in [129, 179]. However, they are non-constant and context specific, so selecting an appropriate elasticity to apply can be perilous. Erhardt et al. [238] explore the range of elasticities encountered for different toll road scenarios as the pricing scheme and level of competition vary. TRL Limited provides detailed analysis and guidance on the use of elasticities as they relate to transit demand [228], and elasticities are a core component of the application of Transit Cooperative Research Program (TCRP) Report [39]. Both acknowledge the limitations of such an approach, but the attention given to the topic makes it clear just how common it is.

Of greatest interest to this study are the elasticities of transit demand with respect to changes in vehicle kilometres or vehicle miles. For the studies included in TCRP 95, the elasticities of bus ridership with respect to a change in bus service range from 0.33 to 1.34. The meta-analysis included in the TRL report shows the range to be 0.10 to 0.74 in the short term, and 0.22 to 1.04 in the long term. The evidence as it relates to rail is more sparse. TCRP 95 found a single study for rail, based on the London Underground, which found the elasticity with respect to a change in service to be 0.08, or about half the value for London buses [239]. TRL's meta-analysis included three rail studies, and found the short run elasticities to range from 0.65 to 0.90.

Wardman [240] provides a more extensive review and meta-analysis of time elasticities of travel demand in the UK. The focus is on travel time, rather than service miles, but there is evidence provided on headway elasticities, which is of interest to us, because many of the changes we consider are service mile changes that result in a change in headway, rather than an extension (the BART to SFO extension being a notable exception). Wardman finds the average headway elasticity to be -0.29 for bus and -0.26 for train. Graham et al [241] derived a long-run elasticity of metro ridership with respect to changes in service frequency of (measured as rail car kilometres per route kilometre) of 0.51. Litman [242] also assembles a very extensive set of travel demand elasticities from a number of studies. He reports a range of 0.6 to 1.0 for the elasticity of transit demand with respect to changes in service miles. A number of other examples can be found, but for the purpose of this analysis, we need a reasonable range for comparison, and accept these as a point of comparison.

Fare elasticities are also relevant to one of the examples. TCRP 95 finds a range of published transit fare elasticities from -0.1 to -0.6, with a "common rule" value of -0.3. TRL reports the overall bus fare elasticity as -0.41 and the overall metro fare elasticity as -0.29, both reported as short run values. Wardman [243] provides a very extensive review and meta-analysis of price elasticities for surface travel demand in the UK. He finds that the average price elasticity of the studies reviewed is -0.46 for bus, -0.28 for underground and -0.86 for rail. The bus and underground values are comparable to TRL's findings, with a higher rail elasticity. For our purposes, we consider BART to be more comparable to underground than to rail, because the travel market is intra-urban, the rider experience is more comparable and the fare structure is similar. Litman also reports that urban rail (such as Chicago's L) fare elasticities are often lower than bus fare elasticities.

The final option considered is to estimate and apply regression models using time series data. This allows the project effect to be estimated while controlling for factors, such as the change in employment, that may also affect demand. In this research, the preferred models from Chapter 4 are used as the starting point, which take the form of Regression with ARIMA Errors (RegARIMA) models. The models are applied to evaluate the change using three possible approaches. One simply applies the change in the value of service variables in combination with the existing model coefficients to estimate the change in demand. The other two consider that the specific service variables are likely to be a coarse representation of the actual service change, and estimate a constant associated with the change to capture the effect not fully explained by a variable such as service miles. The two differ in whether all variables are re-estimated or only the constant.

6.1.1 2010 MUNI Service Cuts

In May 2010, MUNI implemented system-wide service cuts in an effort to close a budget deficit [185]. The cuts reduced transit frequencies, particularly in the late night period, and amounted to an approximately 10% service reduction. Service was partially restored in September 2010. Related work used the data produced by the data fusion tool developed in Chapter 3 to explore the implications of these cuts by comparing specific months before and after the cuts were implemented [117, 244]. Here, the time series models estimated in Chapter 4 are used to estimate the ridership effect of those cuts, which is compared to the estimated ridership effect using before-and-after comparisons and using an elasticity calculation.

Table 6.1 shows three RegARIMA models estimated to capture the effects of these service cuts, which went into effect on 8 May 2010. The base model is identical to the preferred model estimated in Chapter 4. The constrained and re-estimated models follow the same specification, but add one additional variable: a constant for the service change. This constant is set to one for each month starting from May 2010, and zero for each month prior to May 2010. In the constrained model, the regression coefficients previously included in the model are fixed, and only the constant and the ARIMA terms are re-estimated. In the re-estimated model, all coefficients are re-estimated.

The column labelled "Change in Value" shows the change in the value of model variables associated with the service cuts. The cuts are associated with a reduction of 5,582 weekday bus service miles and 2,984 weekday MUNI rail service miles. The constant for the service change switches from zero to one. The final row of the table shows the estimated effect of the service change. This is calculated as the product of the change in variable values times the model coefficients.

Table 6.1: Estimated effect of MUNI 2010 service cuts

Model Characteristics								
Dependent variable	MUNI boardings							
Type	$ARIMA(0,1,1)(0,1,0)_{12}$							
Estimation date range	Jun 2009 to Nov 2013							
Service change date	8 May 2010							
Predictive Variables								
		Change in	Base Model		Constrained		Re-estimated	
Description	Lag	Value	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1		-0.3092	-1.67	0.2991	-1.80	-0.2938	-1.52
Weekday service miles, 1000s		-5.582	7971	2.57	7971	fixed	7347	1.84
Weekday service miles on MUNI rail, 1000s		-2.984	-2777	-1.12	-2777	fixed	-2800	-1.12
Average bus runspeed		-0.34	49853	1.94	49853	fixed	49540	1.94
Employment in San Francisco			0.876	fixed	0.876	fixed	0.876	fixed
Constant for service change		1			-3319	-0.17	-7247	-0.25
Model Statistics								
RMSE			15,401		15,397		15,391	
Percent RMSE			3.09%		3.09%		3.09%	
Box-Pierce test p-value			0.873		0.820		0.808	
Effect of Service Change								
Estimated total effect			-53,158		-56,477		-56,748	

The results of the constrained model estimation show a coefficient of -3,419 on the service change constant. This implies that the actual change in ridership is slightly worse than would be implied simply by the change in service miles on bus and rail. It is logical to expect that the exact schedule, routes cut, and other details would affect the ridership, so it is reasonable that there may be some non-zero constant. The t-statistic for this constant is only -0.17, though, indicating that it is not significantly different from zero. The remaining regression coefficients are fixed, and the moving average coefficient only changes slightly.

When the full model is re-estimated, the service change constant is higher, at -7,247, although the t-statistic indicates that it is still not significantly different from zero. This higher service change constant is offset by a lower coefficient on weekday service miles. Both are changing at the same time, so by capturing some of the effect in the constant, the service miles coefficient is able to adjust so it better matches changes at other points in time.

When the change in variable values are applied to the base model, the results indicate that the service change is associated with a reduction of 53,200 riders for an average weekday. When applied with the constrained model, the estimated effect is a reduction of 56,500 riders. The difference between the two is the inclusion of the constant. The re-estimated model implies a reduction of 56,700 riders. These three approaches provide a relatively narrow range of estimated effects, so it appears that it may be reasonable to use any, without a large risk of fundamentally changing our understanding of the effect of the change. These estimated changes represent an approximately 12% reduction in ridership associated with a 10% service cut.

As a point of comparison, Table 6.2 shows a comparison of ridership before and after the May 2010 service cuts. To avoid capturing seasonal effects, the year-over-year difference is shown for each of 10 months before the cuts, and each of 10 months after the cuts. Depending on which month is selected, the difference ranges from -44,000 riders to +6,000 riders. The average over the 10 month period is -23,000. The variation in the difference by month highlights how challenging it can be to report meaningful before-and-after results.

Next, elasticities are used to calculate the change in ridership associated with these service cuts. Table 6.3 reports the results of these calculations. The published elasticities of bus ridership with respect to service miles range from 0.10 to

Table 6.2: Before and after calculations for MUNI 2010 service cuts

Date		MUNI Boardings		
Before	After	Before	After	Difference
Jun-09	Jun-10	522,835	485,692	-37,144
Jul-09	Jul-10	510,301	491,666	-18,635
Aug-09	Aug-10	514,248	493,783	-20,465
Sep-09	Sep-10	547,166	517,500	-29,666
Oct-09	Oct-10	528,611	519,322	-9,290
Nov-09	Nov-10	524,436	494,313	-30,123
Dec-09	Dec-10	467,521	458,600	-8,920
Jan-10	Jan-11	484,088	490,064	5,976
Feb-10	Feb-11	529,754	489,776	-39,979
Mar-10	Mar-11	533,800	489,976	-43,824
10 Month Average		516,276	493,069	-23,207

1.34 [228, 39], implying a ridership change from -5,000 to -67,000. The before-and-after results (Table 6.2) fall within this range for all months, and the estimated change (Table 6.1) is at the high end of this range. This is a wide range, so it may not be worth bragging about the model's ability to fall within this range.

Table 6.3: Elasticity calculations for MUNI 2010 service cuts

Starting Ridership	514,248	
	Low	High
Percent change in service miles	-9.67%	
Published elasticity	0.10	1.34
Calculated ridership change	-4,971	-66,605

The corollary question becomes: if the RegARIMA model predicts a different effect than simple before-and-after tabulations, what accounts for the difference? This question is examined in a manner similar to the analysis of the factors contributing to changes in ridership presented in Chapter 5.

Table 6.4 summarises the factors contributing to the change in ridership between August 2009 and August 2010. The August to August change was selected because its change is most similar to the average change over the 10 month period. The re-estimated model from Table 6.1 is used to calculate these changes. The change in weekday service miles and the change in weekday service miles on MUNI rail are the two factors associated with the May 2010 service cuts. The other factors are unrelated. The analysis shows that the change associated with the May 2010 service cuts is -64,000, but that this is partially offset by an increase of 28,000 riders associated with other explained factors in the model, and an increase of 17,000 associated with an unexplained trend and the random variation. This analysis suggests that if not for the service cuts, MUNI ridership would have increased. The total change in ridership between the two months is -20,000.

Table 6.4: Contributions to change in MUNI ridership: Aug 2009 to Aug 2010

			Value		Ridership Change	
Description	Lag	Coef	Aug-09	Aug-10	Absolute	Percent
Weekday service miles, 1000s		7,347	57.75	52.05	-41,921	-8.2%
Weekday service miles on MUNI rail, 1000s		-2,800	16.71	13.28	9,616	1.9%
Average bus runspeed		49,540	10.94	10.60	-16,844	-3.3%
Employment in San Francisco		1	541,409	548,146	5,902	1.1%
Constant for service change		-7,247	0	1	-7,247	-1.4%
Unexplained Trend			-879,378	-844,874	34,504	6.7%
Residual			-144	-4,619	-4,475	-0.9%
Associated with May 2010 service cuts					-56,396	-11.0%
Associated with other explained factors					5,902	1.1%
Associated with unexplained trend or random variation					30,029	5.8%
Total Ridership			514,248	493,783	-20,465	-4.0%

6.1.2 2009 BART Service Cuts

In 2009 BART faced a budget deficit [245], and chose to meet that deficit with a combination of a fare increase and service cut [246]. The fare increase was a 6.1% adjustment implemented on July 1, 2010. The service cuts were a decrease in train frequency, amounting to approximately 5% less overall service, implemented on September 14, 2009.

Table 6.5 shows three RegARIMA models estimated to capture the effects of the 2009 BART service cuts and fare increase. The base model is identical to that estimated in Chapter 4, while the constrained and re-estimated models include a constant for the service change. Because the cuts take effect halfway through the month, the value of the constant term is set to 0.5 for September 2009, one for all months after September 2009, and zero for all months prior to September 2009. The constant is specified with a 12-month distributed lag, as defined in Section 4.3.6.

This means that 1/12th of the coefficient value is added each month, up to a maximum of 12 months. It is consistent with the lag in changes to service miles and stations used in the broader model estimation.

For the constrained model, the estimated service change constant is -867, indicating that the ridership effect of these cuts is slightly worse than would be expected purely from the service change. This constant is not significantly different from zero, however. The re-estimated model produces a larger constant, offset by a somewhat smaller service miles coefficient. The constant in the re-estimated model is still not significant.

The change in values associated with these service cuts are a reduction of 1,531 service miles, fare increase of $0.22, and the service change constant. The estimated total effect of these service changes ranges from -8,700 riders for the base model to -9,900 riders for the re-estimated model. For a 5% cut in service miles and a 6% increase in fares, BART ridership is reduced by between 2.5% and 2.9%. This is a lesser effect than shown for MUNI, which may be because there is less competition of other modes with BART. It is consistent with previous evidence from London of the relative elasticities of bus and rail ridership with respect to service changes [46].

Table 6.5 shows a comparison of BART ridership before and after the September 2009 service cuts. Here, the values in each month are reported for the 10 month period prior to the change, compared to the 10 month period starting on year after the changes. This delay allows the lagged effect to be incorporated. This table shows that the before-and-after difference ranges from -25,000 riders to +16,000 riders, with an average change of -6,200 riders. The broad range of changes makes it more challenging to look at these data and determine what the most appropriate estimate of the change might be.

Table 6.7 shows the application of elasticities to calculate the ridership effect of these 2009 BART service cuts. The published elasticities for urban rail ridership with respect to changes in service miles range from 0.08 to 0.90 [228, 39]. These elasticities correspond to a ridership change of between -1,500 and -16,700. The published fare elasticities range from -0.10 to -0.60 [228, 39]. The total rider-

Table 6.5: Estimated effect of BART 2009 service cuts

Model Characteristics								
Dependent variable	BART boardings							
Type	$ARIMA(0,1,2)(0,1,1)_{12}$							
Estimation date range	Jan 2001 to Mar 2015							
Service change date	14 Sep 2009							
Predictive Variables								
		Change in	Base Model		Constrained		Re-estimated	
Description	Lag	Value	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1		-0.570	-5.08	-0.571	-7.23	-0.579	-4.57
Moving average coefficient	2		-0.283	-2.74	-0.281	-3.34	-0.285	-2.62
Seasonal moving average coeff.	S1		-0.660	-8.44	-0.658	-8.96	-0.659	-8.34
Weekday service miles, 1000s	D(0,12)	-1.531	2712	2.07	2712	fixed	2400	1.53
Number of Stations	D(0,12)		5472	5.18	5472	fixed	5515	5.31
Employment in 4-county area			0.2027	10.96	0.2027	fixed	0.2005	9.69
Percent of 4-county employment in SF			8099	2.10	8099	fixed	8496	1.83
Cash fare (2010 $)		0.22	-20795	-2.50	-20795	fixed	-20875	-2.41
Average car fuel cost (2010 $/mile)			86312	2.74	86312	fixed	89439	2.70
Days with a BART strike			-19010	-21.0	-19010	fixed	-18999	-20.80
Constant for service change	D(0,12)	1			-867	-0.24	-1709	-0.35
Model Statistics								
RMSE			4923		4923		4920	
Percent RMSE			1.42%		1.42%		1.42%	
Box-Pierce test p-value			0.207		0.162		0.154	
Effect of Service Change								
Estimated total effect			-8,685		-9,552		-9,934	

Table 6.6: Before and after calculations for BART 2009 service cuts and fare increase

Date		BART Boardings		
Before	After	Before	After	Difference
Sep-08	Sep-10	379,996	354,579	-25,417
Oct-08	Oct-10	370,502	356,545	-13,957
Nov-08	Nov-10	364,329	353,502	-10,827
Dec-08	Dec-10	351,130	323,669	-27,461
Jan-09	Jan-11	344,786	334,045	-10,741
Feb-09	Feb-11	348,046	346,196	-1,850
Mar-09	Mar-11	346,658	343,971	-2,687
Apr-09	Apr-11	347,237	352,160	4,923
May-09	May-11	342,394	352,052	9,658
Jun-09	Jun-11	339,133	355,436	16,303
12 Month Average		353,421	347,215	-6,205

ship effect reflects the combination of both. The estimated effects in Table 6.5 fall towards the low to middle part of this range. The average before-and-after calculations shown in Table 6.5 are towards the lower end of this range.

Table 6.7: Elasticity calculations for BART 2009 service cuts and fare increase

Starting Ridership	348,046	
	Low	High
Percent change in service miles	-5.33%	
Published service elasticity	0.08	0.90
Calculated ridership change	-1,485	-16,706
Percent change in fares	6.10%	
Published fare elasticity	-0.10	-0.60
Calculated ridership change	-2,123	-12,738
Total ridership change	-3,608	-29,444

Table 6.8 shows the factors contributing to the change in BART ridership between March 2009 and March 2011. March was selected as the month closest in value to the 10-month average. This analysis is based on the re-estimated model, shown in Table 6.5. It indicates that the ridership change associated with the service cuts is -9,900. Because the fare values are adjusted for inflation, the inflation effect is reported separately from the July 2009 fare increase. In total, the fare increase is only slightly larger than the effect of inflation over the two year period. Explained factors, beyond the service cut and fare increase, contribute to a relative increase in ridership, offsetting about half the policy effect. The unexplained trend and residual contribute towards some additional relative increase. The total effect is a reduction of 4,000 BART riders.

6.1.3 2003 BART SFO Extension

After a planning process that dates to the original inception of the system in the 1960s, BART opened an 8.7 mile extension to SFO in June 2003 [247]. In addition to a station at SFO, it included three additional stations in San Mateo County, just south of San Francisco. Two years into its operation, ridership on the extension was lower than forecast, and operational adjustments were made in an effort to contain costs [248]. A few studies have examined conducted retrospective evaluations of this extension in varying degrees of detail. Shortly after its opening, West

Table 6.8: Contributions to change in BART ridership: Feb 2009 to Feb 2011

			Value		Ridership Change	
Description	Lag	Coefficient	Mar-09	Mar-11	Absolute	Percent
Weekday service miles, 1000s	D(0,12)	2400	28.71	27.17	-3,692	-1.1%
Number of Stations	D(0,12)	5515	43.00	43.29	1,626	0.5%
Employment in 4-county area		0.2005	1,865,020	1,812,752	-10,480	-3.0%
Percent of 4-county employment in SF		8496	29.86	30.28	3,593	1.0%
Cash fare (2010$): Inflation before change		-20875	3.42	3.37	1,159	0.3%
Cash fare (2010$): Fare change		-20875	3.37	3.58	-4,550	-1.3%
Cash fare (2010$): Inflation after change		-20875	3.58	3.45	2,718	0.8%
Average car fuel cost (2010 $/mile)		89439	0.10	0.17	6,335	1.8%
Days with a BART strike		-18999	0	0	0	0.0%
Constant for service change	D(0,12)	-1709	0	1	-1,709	-0.5%
Unexplained Trend			-521,516	-519,410	2,106	0.6%
Residual			-1,519	-2,697	-1,178	-0.3%
Associated with Sep 2009 service cuts					-9,951	-2.9%
Associated with other explained factors					4,951	1.4%
Associated with unexplained trend or random variation					927	0.3%
Total Ridership			348,046	343,971	-4,073	-1.2%

and Herhold [249] compared forecasts from various stages of planning to actual ridership. They found that in the first month after opening, ridership specifically at the SFO BART station "mostly matched recent expectations." These levels were not maintained, however, as subsequent research reported a drop in ridership by December 2003 [250]. That latter analysis also focused specifically on the SFO station. It compared the air passenger mode shares using BART to those of other comparable systems, and found them to be in line with what might be expected. Freeman et al. [247] and Mason [251] each examined the planning process that led to the selected design and found that the complexity of that planning process and the trade-offs made in the political process led to a design that may have been less than ideal. Together, these two studies provide an excellent overview of the history of the project and the factors involved in the decision making process.

This analysis adds to that existing knowledge base by applying the RegARIMA models developed in Chapter 4 to estimate the ridership increase associated with the extension, and distinguish it from other background factors that are changing over this period. In doing so, it provides an opportunity to validate the time series models by comparing the predicted ridership difference to the entries and exits specifically at the four new stations.

Table 6.9 shows the three RegARIMA models estimated to capture the ridership effect of the BART extension to SFO. The extension is associated with four new stations and an increase of 1,517 service miles, as shown in the change in value column. The constrained model shows a positive, but insignificant constant associated with the service change. The re-estimated model shows a larger positive constant, although it is still not significant. There is an associated decrease in the size of the service miles and number of stations coefficients in the re-estimated model. As with the models above, the constant is specified with a 12 month distributed lag.

These model results indicate that the ridership increase associated with the SFO extension is between 26,000 and 27,000 riders. This is approximately a 9% ridership increase associated with a 6% increase in service miles and a 10% increase in the number of stations.

Table 6.10 shows the BART ridership before and after the SFO extension. To account for lagged effects, the after period does not begin until 12 months after the opening of the extension. These data show that the increase ranges from 6,000 riders to 27,000 riders, depending on the month. The average change over the 12 month periods is 14,000 riders. February is the month closest to the 12-month average.

As a second point of reference, the station level passenger entries and exits were tabulated for the new stations [143]. These data show 20,500 entries and exits on the new stations for an average weekday in February 2005. By February 2010, there were 28,400 entries and exits at these four stations, and by February 2015 it they had increased to 30,500. The value predicted by the RegARIMA models is larger than the 2005 station-specific ridership, and slightly lower than the 2010 station-specific ridership. The models are specified with additive variables, which constrains the estimated ridership change to a single value. As a future test, it would be interesting to test whether a multiplicative term does a better job of capturing the ridership trend on the SFO line.

All of these values, however, are much lower than the forecasts included in the

Table 6.9: Estimated effect of BART SFO extension

Model Characteristics								
Dependent variable	BART boardings							
Type	$ARIMA(0,1,2)(0,1,1)_{12}$							
Estimation date range	Jan 2001 to Mar 2015							
Service change date	22 Jun 2003							
Predictive Variables								
		Change in	Base Model		Constrained		Re-estimated	
Description	Lag	Value	Coef	T-Stat	Coef	T-Stat	Coef	T-Stat
Moving average coefficient	1		-0.5701	-5.08	-0.569	-7.20	-0.563	-5.51
Moving average coefficient	2		-0.2827	-2.74	-0.278	-3.22	-0.267	-2.74
Seasonal moving average coeff.	S1		-0.6603	-8.44	-0.658	-8.95	-0.658	-8.53
Weekday service miles, 1000s	D(0,12)	1.517	2712	2.07	2712	fixed	2358	1.64
Number of Stations	D(0,12)	4	5472	5.18	5472	fixed	3176	1.08
Employment in 4-county area			0.2027	10.96	0.2027	fixed	0.2076	10.87
Percent of 4-county employment in SF			8099	2.10	8099	fixed	7663	2.25
Cash fare (2010 $)			-20795	-2.50	-20795	fixed	-22673	-2.66
Average car fuel cost (2010 $/mile)			86312	2.74	86312	fixed	80085	2.47
Days with a BART strike			-19010	-21.0	-19010	fixed	-19055	-21.26
Constant for service change	D(0,12)	1			1074	0.27	11264	0.85
Model Statistics								
RMSE			4923		4923		4915	
Percent RMSE			1.42%		1.42%		1.42%	
Box-Pierce test p-value			0.207		0.169		0.187	
Effect of Service Change								
Estimated total effect			26,002		27,076		27,544	

final 1996 environmental impact statement (EIS) for the project, which projected 69,000 daily riders by 2010 [249]. To be fair, the forecasters during the booming economy of the 1990s would not have expected 2010 Bay Area employment to be only marginally higher than its 1996 level, and nearly 10% lower than in 2000 [210], but this only serves to further highlight the challenges of long-term forecasting for major infrastructure projects.

Published elasticities were used to calculate the ridership effect of the BART extension to SFO, using the same elasticities as in the previous sections. Specific elasticities were not available for urban rail extensions, so the increase in service miles was used instead, although it is admittedly a crude proxy. Table 6.11 shows the result, which indicates a ridership increase of between 1,400 and 15,800 passengers. If the same elasticities were applied to the 10% increase in the number

Table 6.10: Before and after calculations for BART SFO extension

Date		BART Boardings		
Before	After	Before	After	Difference
Jun-02	Jun-04	302,730	308,793	6,063
Jul-02	Jul-04	300,921	308,190	7,269
Aug-02	Aug-04	296,473	304,724	8,251
Sep-02	Sep-04	309,640	323,236	13,596
Oct-02	Oct-04	303,796	315,753	11,957
Nov-02	Nov-04	294,360	313,975	19,615
Dec-02	Dec-04	282,622	294,788	12,166
Jan-03	Jan-05	288,638	297,524	8,886
Feb-03	Feb-05	295,525	310,053	14,528
Mar-03	Mar-05	296,030	310,722	14,692
Apr-03	Apr-05	292,949	319,649	26,700
May-03	May-05	289,681	315,888	26,207
12 Month Average		296,114	310,275	14,161

of BART stations, the estimated ridership increase would be between 2,400 and 27,000 passengers. The monthly before-and-after differences generally fall within these ranges, while the estimated effect in Table 6.9 is at the high end of the latter range. It is logical that constructing new track would have a larger effect than simply adding service to existing track.

Table 6.11: Elasticity calculations for BART SFO extension

Starting Ridership	295,525	
	Low	High
Percent change in service miles	5.94%	
Published elasticity	0.08	0.90
Calculated ridership change	1,404	15,792

Table 6.12 shows the factors contributing to the change in BART ridership between February 2003 and February 2005. The analysis uses the re-estimated model from Table 6.9. The change stations and the constant are associated with the SFO extension. The service miles term is broken into two components. When the SFO extension opened in 2003, the BART service miles increased from 25,550 to 27,090. Service was later cut, in 2004, back to 26,190 service miles on an average weekday. For the purpose of this analysis, the 2003 increase is treated as a part of the project, and the 2004 cut is treated an unrelated measure.

The analysis shows that the increase associated with the SFO extension is 28,000 riders, while there is a decrease of 14,000 associated with other explained factors. About 60% of the drag on ridership is associated with a decline in employment and in the share of employment in San Francisco. The other 40% is associated with factors within BART's control: the service cuts in 2004 and a fare increase also in 2004.

Table 6.12: Contributions to change in BART ridership: Feb 2003 to Feb 2005

			Value		Ridership Change	
Description	Lag	Coefficient	Feb-02	Feb-05	Absolute	Percent
Weekday service miles, 1000s: SFO extension	D(0,12)	2358	25.55	27.09	3,636	1.2%
Weekday service miles, 1000s: Other cuts	D(0,12)	2358	27.09	26.19	-2,116	-0.7%
Number of Stations	D(0,12)	3176	39.00	43.00	12,702	4.3%
Employment in 4-county area		0.2076	1,876,808	1,845,497	-6,500	-2.2%
Percent of 4-county employment in SF		7663	28.26	28.01	-1,915	-0.6%
Cash fare (2010 $)		-22673	3.29	3.46	-3,837	-1.3%
Average car fuel cost (2010 $/mile)		80085	0.11	0.11	198	0.1%
Days with a BART strike		-19055	0	0	0	0.0%
Constant for service change	D(0,12)	11264	0	1	11,264	3.8%
Unexplained Trend			-426,070	-427,950	-1,881	-0.6%
Residual			-2,559	418	2,977	1.0%
Associated with SFO extension					27,602	9.3%
Associated with other explained factors					-14,170	-4.8%
Associated with unexplained trend or random variation					1,096	0.4%
Total Ridership			295,525	310,053	14,528	4.9%

6.2 Forecasting Applications

In this section, two examples are considered to demonstrate how the time series models estimated in Chapter 4 can be applied to generate short-term forecasts that may assist in establishing reference points against which projects can be evaluated. Such applications are of particular importance as performance-based planning becomes more common.

In performance based planning, planning goals and objectives are aligned with specific performance measures against which projects and policies can be evaluated. A range of examples are provided in [151]. Recent federal transportation legislation establishes performance based planning as central to the broader planning process [32].

The San Francisco Municipal Transportation Agency (SFMTA) strategic plan [252] is a good example of performance based planning. SFMTA oversees transport in San Francisco, and in that role it both operates the MUNI transit system, and is responsible for parking and traffic. The strategic plan reflects this combined role, and sets four strategic goals:

1. Create a safer transportation experience for everyone.
2. Make transit, walking, bicycling, taxi, ridesharing, and carsharing the preferred means of travel.
3. Improve the environment and quality of life in San Francisco.
4. Create a workplace that delivers outstanding service.

Each goal is associated with a set of performance indicators and associated targets, which are regularly monitored and reported [253]. For goal two, which is of the most relevance here, the performance indicators and targets are:

1. Improve the customer satisfaction rating by 0.5 points for each budget cycle.
2. Eliminate transit bunches and gaps for 25% of ridership.
3. Reduce the private auto mode split to below 50%.
4. Maintain a 75% to 85% occupancy of public metered parking spaces in the areas managed by the SFPark system.

The mobility goals do not explicitly consider transit ridership. Instead, it is implicit in the broader goal of reducing the private auto mode split. Given that recent trends in MUNI ridership (observed and discussed in Chapter 5) show an unexplained downward trend in MUNI ridership, but growth in commutes by walk, taxi, bike and other modes, this is perhaps both a logical and an achievable goal.

More broadly, this highlights a key challenge of performance based planning: that a range of factors beyond the control of the planning agency will affect whether or not a target is met. It is possible that either pre-existing trends can be used as a basis for declaring success for an intervention, or that trends in the opposite direction can undermine the perception of what is otherwise a successful intervention. Instead, the goal should be to set targets that meaningfully inform or promote the development of a better transport system. Because the planning process implies that targets should be set a priori, short term forecasts of performance with no intervention can be a valuable tool to help set targets. This is further illustrated at the project level with two examples of pilot projects in the Bay Area.

6.2.1 MUNI 5L Fulton Pilot Project

This example considers a route-level service change rather than a major set of system-level changes. In this example, MUNI implemented a pilot project for the 5L Fulton line aimed at increasing bus speed and improving ridership. One goal of the pilot project was to increase ridership on the 5L by a fixed percentage. The analysis presented here considers the the rate of growth in MUNI ridership that could be expected systemwide, and compares that growth to the target ridership increase for the pilot.

In October 2013, the San Francisco Municipal Transportation Agency (SFMTA) opened the 5L Fulton pilot project [254]. Operating in a busy east-west transit corridor traversing the city, the project seeks to reduce crowding, reduce travel times and improve reliability. As shown in Figure 6.1 the core change is that the existing 5 route is spilt into two parallel bus routes. The 5 makes all stops, but only runs for a portion of the corridor, while the new 5L runs the full length of the corridor, but skips some stops where it overlaps with the 5. Thus, the 5L is designed to offer faster service to those who travel further. In addition a series of complementary physical changes are made in the corridor, including the removal of some bus stops altogether, changes to the location of other stops relative to signals and reducing the number of travel lanes in a section of the corridor.

The 5L is important because it serves as a test for similar changes that are planned throughout the city as part of the Transit Effectiveness Project [255]. Prior to the pilot launch, SFMTA established an evaluation metric for the project, related to ridership [256]. The metric was defined as:

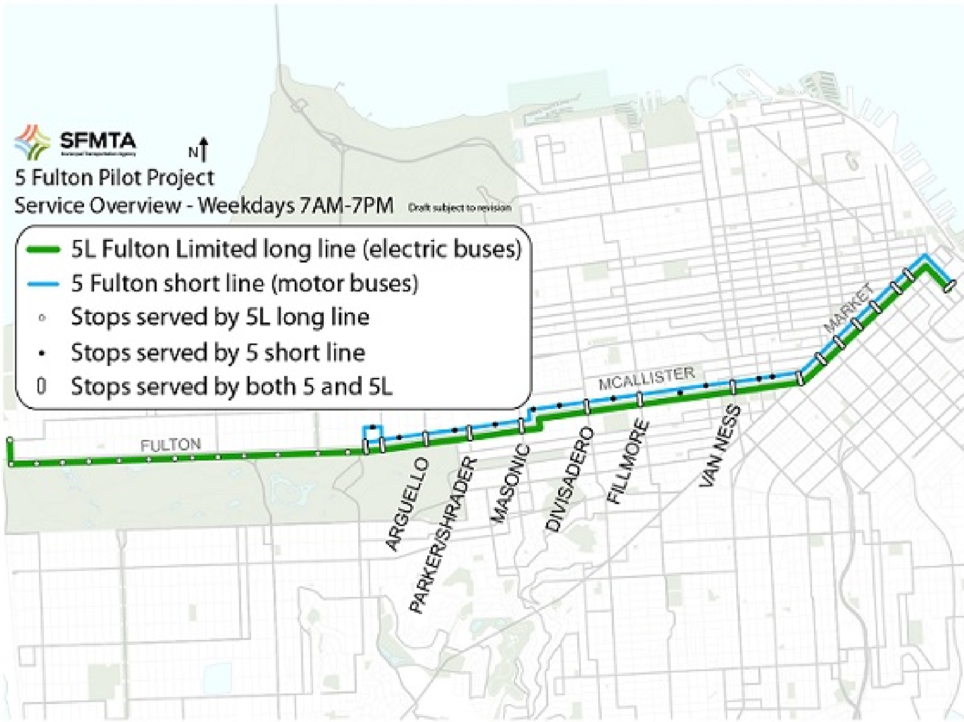


Figure 6.1: Overview of service for 5L Fulton pilot project [254]

- Ideal: Beyond six months after pilot implementation, when compared with observations made prior to pilot implementation, ridership increases by 3%.
- Meets Standard: Beyond six months after pilot implementation, ridership increases by less than 3%.
- Substandard: Beyond six months after pilot implementation, when compared with observations made prior to pilot implementation, ridership does not increase.

In establishing such metrics, it would be useful to put them in the context of the ridership change that would otherwise be expected. To do this, the preferred MUNI RegARIMA models (from Table 4.15) were applied to forecast the MUNI bus ridership from December 2013 through March 2015. If more detailed results were desired, a similar set of forecasts could be calculated for a time series specific to the 5/5L route. For these forecasts, the observed employment in San Francisco County is used, taken from the Quarterly Census of Employment and Wages (QCEW). The level-of-service characteristics (service miles, rail service miles and runspeed) are extended from the last value of the observed time series, assuming no change.

Figure 6.2 shows a plot of the observed time series in black, and the forecast time series in blue. Table 6.13 shows a comparison of the modelled MUNI ridership in Autumn 2012 through Winter 2013, compared to the modelled ridership in Autumn 2014 through Winter 2015. The model predicts that on average, the ridership in the latter period is 3.9% lower than the earlier period. This indicates that for the pilot project to meet the standard of a ridership increase greater than 0% and less than 3%, the 5 and 5L would need to outperform what is expected from the system as a whole.

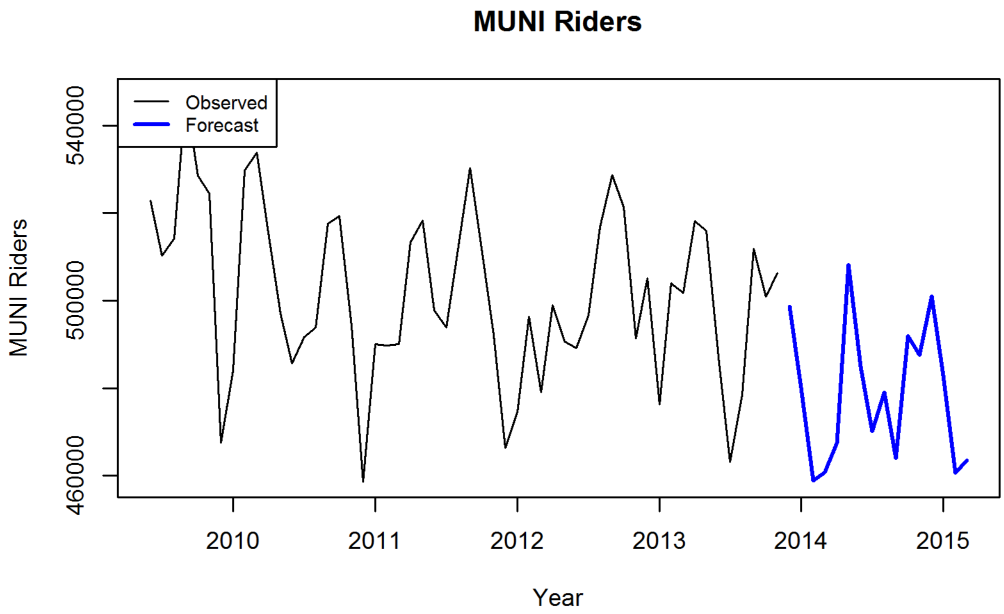


Figure 6.2: MUNI ridership forecast

One caveat to this analysis is that there is a question as to what the best point of comparison is for ridership trends on a particular route. In other words, what is the control group? Graham [134] considers the issue of control groups and causal inference in transport. He notes that transport interventions tend not to be randomly assigned (unless you think very poorly of the political decision making process involved), so the evaluation is different than in fields like medicine where randomised control trials are the norm. The selection of a control group is ultimately a matter

Table 6.13: Baseline forecast of MUNI ridership

Date		MUNI Boardings			
Before	After	Before	After	Difference	Pct Diff
Sep-12	Sep-14	523,525	464,049	-59,475	-11.4%
Oct-12	Oct-14	513,029	491,988	-21,041	-4.1%
Nov-12	Nov-14	496,804	487,496	-9,307	-1.9%
Dec-12	Dec-14	475,040	501,018	25,979	5.5%
Jan-13	Jan-15	500,984	482,895	-18,089	-3.6%
Feb-13	Feb-15	494,449	460,621	-33,828	-6.8%
6 Month Average		500,638	481,345	-19,294	-3.9%

of judgement. In this situation, we are treating the system-level trends as a control, which is an improvement over having no control. However, it may be preferable to instead identity specific routes that have similar characteristics to the 5 and 5L: serving downtown San Francisco from a residential area with similar levels of growth and development. One option is the parallel routes, as discussed in the next paragraph. While comparable in many ways, these have their own limitation in that they are not independent: improvements to the 5 and 5L may cause some diversion from parallel routes. Acknowledging these limitations, we move forward to consider the actual changes.

In reality, on weekdays when both are in service, ridership on the 5 and 5L increased by 17% between the period from Autumn 2012 through Winter 2013 to Autumn 2014 through Winter 2015 [256]. Ridership on two parallel routes, the 21 Hayes and 31 Balboa, decreased by 6% and increased by 1%, respectively. This change clearly meets the ideal ridership metric, whether the reference point is no change or a 3.9% decrease.

6.2.2 BART Perks Pilot Project

As ridership has increased in recent years, BART patrons face heavily crowded trains, particularly on transbay trains with passengers alighting in downtown San Francisco [257]. A major capacity expansion would involve constructing a second transbay tube, at a cost of up to $12 billion [205]. In an effort to explore lower cost interventions to mitigate the crowding issues, in March 2016, BART launched a six-month pilot project called BART Perks [258]. Administered by the Bay Area tech startup Urban Engines [259], BART Perks involves recruiting BART commuters to sign up for the program in which they will download a smartphone app, and link their Clipper Card to their profile. They will then be offered incentives to shift their commute out of the peak period. The incentives scheme that Urban Engines offers is based on research conducted by the founders of Urban Engines at Stanford University [260, 261].

From a planning perspective, there are two questions for which this research is relevant. First, are the biases in Clipper Card use problematic for the project implementation? Second, how will the effectiveness of the pilot project be evaluated?

In terms of the biases, it was observed in Chapter 2 that minorities and low income travellers are less likely to use a Clipper Card as a means of fare payment. For a program that requires the use of a Clipper Card for the purpose of monitoring peak travel, as well as a smartphone app to participate, these biases raise two further questions. First, does it limit the effectiveness of the program to exclude low-income travellers who could reasonably be expected to have a lower value of time and therefore be more responsive to incentives? Second, does it create an equity problem to give incentives exclusively to a population that tends to be higher income and more white? We do not answer these questions here, but merely raise them as issues to be considered.

For the evaluation, it is important to consider how to measure the effectiveness of the pilot in a way that would help BART decide whether or not to continue the program. In January 2016, I engaged with the planners at the San Francisco County Transportation Authority (SFCTA) and their contractor responsible for evaluating the pilot. In early discussions with Urban Engines and BART, several performance metrics had been discussed, including the number of enrollees, and a possible target of a 5% reduction in peak period BART trips to the four most crowded stations. It was noted that the former only incentivised program enrolment and not behavioural change, while the latter could be challenging in the context of overall ridership growth.

To gain further insight into the expected baseline growth, the preferred BART

RegARIMA model was applied to forecast the system-wide BART ridership for a six-month pilot period in 2016, compared to the same period in 2015. It is acknowledged that this forecast is not specific to the stations of interest in the peak period, and does not reflect a capacity constraint. Nonetheless, it does give a broad picture of the expected demand trend.

Starting from the estimation data set that extended to March 2015, the forecast was run for a 24 month period from April 2015 through March 2017. To generate the forecasts, the descriptive variables were first forecast or extended. Observed employment data was available through September 2015 from the Quarterly Census of Employment and Wages (QCEW). Values beyond that point were forecast using a seasonal ARIMA model. Separate models were developed for the total four-county employment, and for employment in San Francisco county. They were combined to calculate the share of employment in San Francisco.

The service miles and number of stations were held constant throughout the forecast period. This excludes the possible effect of the Oakland airport connection and the Warm Springs BART station. The Oakland airport connection is an automated guideway system, operated by BART, that links the Coliseum BART station to Oakland International Airport [262]. It replaces a previous bus route, serving the same connection. It is excluded from this analysis because it is not part of the regular BART system. The Warm Spring Station is a new station that is part of a 5.4 mile BART extension south towards San Jose [263]. It is expected to open in 2016, although the exact date has not been specified [264]. Due to the schedule uncertainty, it is excluded from the analysis. If it were included, the effect in the model would be to shift the ridership forecast upwards by 5,472 riders per station once the full effect is in place.

The fuel cost is extended using the last value of the time series: \$0.14 per mile, in 2010 dollars. It was assumed that the nominal fares would remain constant over the forecast period, so the real value of the fare was scaled assuming 2% annual inflation. No strikes were assumed during the forecast period.

Figure 6.3 shows a plot of the observed time series in black, and the forecast time series in blue. The forecast continues forward the recent trend of strong ridership growth. There is a 12 month period for which observed and forecast values are both available. These are data which were excluded from the model estimation, but recently became available. They provide an opportunity to validate the short-term forecasts.

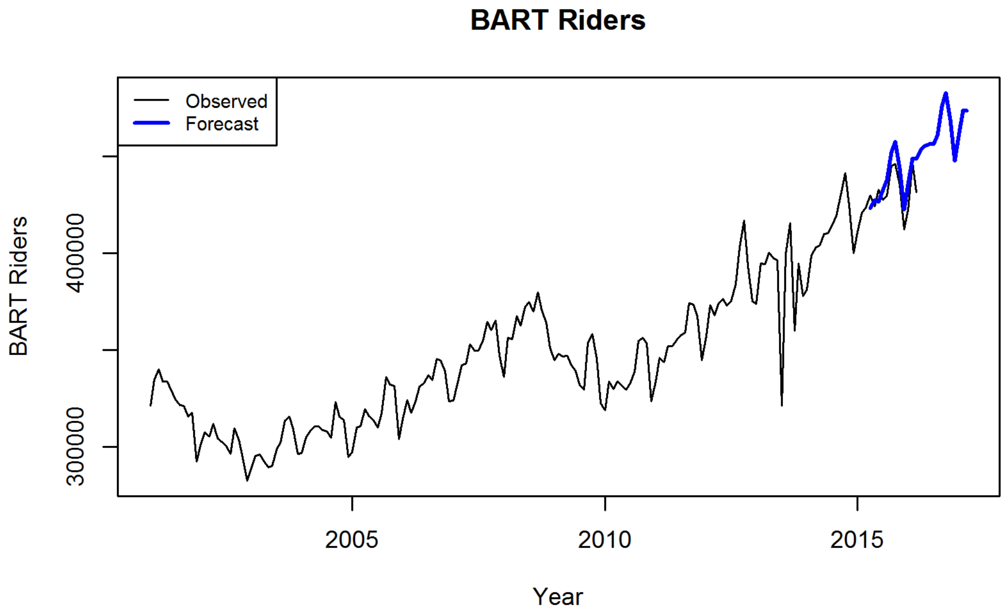


Figure 6.3: BART ridership forecast

Table 6.14 compares the forecast and observed ridership values for this overlapping validation period. The modelled values are, on average, 1.4% higher than observed, with the over-estimates slightly larger towards the latter part of the period. The Root Mean Square Error (RMSE) is 2.1% for these forecasts, which compares to 1.4% RMSE for the estimation period. These results appear reasonable, at least for this relatively short period.

Table 6.14 shows the baseline forecast of BART ridership, comparing the modelled values in April through September 2015 to the modelled values in April through September 2016. The model predicts an average 6.1% growth in ridership for the 2016 period compared to a year earlier. This change is driven dominantly by the projected employment increase following current trends. The projected increase would be even higher if the effect of the Warm Springs BART station were

Table 6.14: Validation forecast of BART ridership

Date	BART Boardings			
Date	Observed	Modelled	Difference	Pct Diff
Apr-15	429,910	423,140	-6,770	-1.6%
May-15	424,312	427,631	3,319	0.8%
Jun-15	432,869	426,717	-6,152	-1.4%
Jul-15	427,467	432,822	5,355	1.3%
Aug-15	429,750	437,768	8,019	1.9%
Sep-15	445,103	452,031	6,928	1.6%
Oct-15	446,008	457,835	11,827	2.7%
Nov-15	435,397	443,893	8,496	2.0%
Dec-15	412,284	422,603	10,319	2.5%
Jan-16	422,314	435,357	13,043	3.1%
Feb-16	446,650	448,911	2,261	0.5%
Mar-16	431,535	448,614	17,079	4.0%
12-month Average	431,966	438,110	6,144	1.4%
RMSE			9,225	2.1%

considered.

Table 6.15: Baseline forecast of BART ridership

Date		BART Boardings			
Before	After	Before	After	Difference	Pct Diff
Apr-15	Apr-16	423,140	453,498	30,358	7.2%
May-15	May-16	427,631	455,268	27,637	6.5%
Jun-15	Jun-16	426,717	456,194	29,477	6.9%
Jul-15	Jul-16	432,822	456,334	23,512	5.4%
Aug-15	Aug-16	437,768	461,099	23,330	5.3%
Sep-15	Sep-16	452,031	475,972	23,941	5.3%
6 Month Average		433,352	459,727	26,376	6.1%

As discussed in the previous case study, the choice of using the system-level trend as the control has limitations. It would be preferable to build a model of the trends at the specific stations in question during the peak period, to account for the possibility that trend at those stations differs from the system as a whole. In addition, this particular case study has one more feature that makes evaluation challenging: a capacity constraint. A time series model estimated from data where the ridership is below capacity may not longer apply once that threshold is met. In such situations, this type of model should be used with caution, particularly for long-term forecasting. Nonetheless, there is value in bringing multiple methods or multiple types of models to the forecasting process, precisely because they may produce different results, and highlight the risks of a project or the limitations of a method.

This analysis highlights the challenge of trying to define a ridership based target to judge the performance of a program such as BART Perks. In the context of 6% projected ridership growth, even a small increase in peak period ridership at the target stations might be considered a successful intervention, depending on the effect of the capacity constraint. Acknowledging these limitations, the planners responsible for the pilot evaluation looked into the feasibility of collecting a control group, such that their behavioural change could be compared to the program participants.

6.3 Conclusions

This research has examined several specific applications of the time series models developed in Chapter 4, using data assembled from the tool described in Chapter 3. The applications fall into two categories: ex-post applications and forecasting applications. The findings from each set of applications are discussed, as well as their potential integration into practice and future research.

6.3.1 Ex-Post Applications

For ex-post applications, the models are applied to three examples of transit service changes to estimate the change in ridership associated with those service changes, controlling for changes in the other terms included in the model. For each example, the model is applied in three ways. First, it is tested using only the service variables included in the original specification: service miles, fare, etc. Then, it is tested with a constant included to capture the effect of the service change beyond what is already captured by the existing variables. When the constant is estimated, the existing coefficients are held constant in the second option, and all coefficients are re-estimated in the third option. For each of the examples tested, the service change constant is not statistically significant, and the difference between the estimated ridership change using each of the three approaches is modest.

The ridership change derived from the time series models is then compared to the ridership change implied by a naive before-and-after difference. In each case, the time series estimate is notably different from the before-and-after estimate. Further, the before-and-after estimates are quite sensitive to the dates selected for the before period and the after period. For example, the average before-and-after difference for the 2009 BART service cuts is -6,200, but the range is from -27,000 to +16,000. The analysis goes on to examine the factors that contribute to the difference between the estimates, and finds that they fall into different categories for each example. This analysis serves to highlight the challenges associated with a simple before-and-after approach.

The time series results are also compared to the ridership change that would be estimated using elasticities found in the literature. For each example, the time series estimates fall within the range of published elasticities, but the range of published elasticities is broad. While useful as a reasonableness check, the elasticity can be highly context specific, and vary depending on the level of competition with other facilities [238]. Therefore, local estimates are preferred to published data where they are available.

6.3.2 Forecasting Applications

Two forecasting applications were considered in the context of performance based planning. Each is for a transit pilot project. In the MUNI 5L Fulton pilot project, a bus route was reconfigured in an effort to improve speed and reliability. In the BART Perks pilot project, an incentive program is being introduced in an effort to shift BART travellers out of the peak period to relieve crowding. In both, there was a desire to set a performance target prior to implementation against which the success of the pilot can be judged. The challenge is similar to that posed by before-and-after studies, because ridership changes can be driven by factors beyond the intervention itself. The time series models predict a downward trend in MUNI ridership and an upward trend in BART ridership over the implementation period, separate from any project effect. These trends can potentially affect the evaluation outcome, and planning agencies would be prudent to consider such short term forecasts when setting targets at the planning stage.

Batty [168] argues that as the nature of available data shifts to include more streaming sources from sensors and related data that are available at a high temporal resolution, there is an accompanying shift towards how we understand cities, with weight added to short term issues. The models presented here further enable that trend, not as a replacement for, but as a complement to long term forecasting models.

6.3.3 Integration into Practice

While the methodology used to analyse these examples is not new (see Washington, Karlaftis and Mannering [202] for a good discussion of such models and their applications) it is clear from the examples studied that when ex-post evaluations are even attempted, they often use very simplistic methods. Hartgen [19] cites a widening gap between research and practice as a challenge to producing good travel forecasts, and the same may be true in ex-post evaluations and in setting performance metrics.

This research has demonstrated that RegARIMA models, estimated from local data, and controlling for a few key factors can offer advantages both for ex-post evaluations and for short term forecasting applications. For ex-post evaluations, they offer an ability to distinguish the project affect from other confounding factors in a way that avoids some of the challenges of a simple before-and after assessment. Short term forecasts from such models can be useful for understanding the baseline trends that may make performance targets easier or harder to hit.

Further, once the framework is in place, the incremental effort needed to apply such models to a range of applications is low. This makes it entirely practical that they can be applied more broadly. The biggest challenge may be in assembling the relevant data. However, the proliferation of continuously collected or regularly updated data sources provides an opportunity to do this. These sources include both transport data, such as automated passenger counters, vehicle location data, transit farecard data, and “drivers of demand” data, such as the Longitudinal Employer-Household Dynamics (LEHD) data. Further work is needed to continue the data wrangling started in Chapter 3 to make these data more accessible, and lower the barriers to using them in models.

6.3.4 Future Research

There are two areas where future enhancements could substantially benefit this research: the integration of highway congestion measures and the estimation of panel data models.

These models notably exclude any measure of traffic congestion, which could reasonably be expected to influence transit ridership. The challenge is that obtaining a good measure of how auto speeds change over time can be difficult. As part of this research, an approach was developed to derive network-wide link speeds from GPS traces on a fleet of taxi cabs, starting from a path inference method developed by Hunter et al [265]. That process could successfully be applied to a few days worth of data, but computational constraints made it impractical for several years' worth of data, at least within the schedule of this thesis. That remains a topic for future development, and integration into the modelling.

It was also observed in Chapter 4 that it can be difficult to obtain strong parameter estimates from the one-dimensional time series models used here. The problem seems to be that there are a limited number of degrees of freedom, because variables change only in time, which makes estimation difficult, particularly when there is colinearity among descriptive variables. One strategy to mitigate this problem would be to switch to panel data models. In this approach, each stop or station would be treated as a separate observation that varies through time, and modelled with variables such as the employment within a certain radius. This is possible with detailed transit observations and the LEHD. It would allow the data to vary both in time and across observations, which may allow for enough variation in the data to obtain stronger parameter estimates. This exercise is left to future research.

Chapter 7

Conclusions

This research developed a data fusion tool that combines several large, continuously collected data sources to monitor travel demand trends, as well as the factors that may influence those trends. It explored the biases and limitations of the data sources used, and estimated time series models from the outputs of that data fusion tool. The planning applications of these models were explored in several examples.

This chapter presents the overall conclusions from that work. It is organised as follows. Section 7.1 reviews the specific findings of the individual research chapters. Section 7.2 proposes several directions for future research. Section 7.3 summarises several lessons learned during the conduct of this research, in an effort to allow future endeavours to benefit from those lessons. Finally, Section 7.4 considers the broader implications for the field of travel forecasting.

7.1 Research Findings

This thesis began with two overarching research questions:

1. How can continuously collected data be leveraged to develop a data fusion tool suitable for monitoring travel demand trends?
2. How can the outputs of that tool be used to gain insight into the drivers of travel demand trends and to measure the transport project impacts?

Rather than seeking to address these questions through broad generalisations, they are explored through specific working examples, as described in Chapters 2 through 6. While they are not the only way to develop such tools and applications, they represent a thoughtful and defensible approach to doing so. By exploring the topic in this way, we have gained insight into the broader questions, with specific conclusions described below.

7.1.1 Transit Smart Card Data Evaluation

Chapter 2 explored the value and limitations of transaction data from the Clipper Card transit smart card system in the San Francisco Bay Area. While transit smart card data have been used for a range of planning applications [87], many of those applications rely on very complete data sets [136, 139, 88, 140]. The Clipper data, however, is subject to strong privacy restrictions, has relatively low penetration rates, and has a high rate of missing data in several important fields. The key question in this situation is: Do the positive aspects of smart card data analysis developed on high quality data sets still apply when the data are more limited? Our findings are as follows.

From a privacy standpoint, it is desirable that a transaction not be precisely located both in space and time. To do so could allow the card holder to be identified by in-person observation. The obfuscation process applied to Clipper data prevents this.

However, this obfuscation relates, in part, to an important limitation of the Clipper data. Except for transactions at fixed rail stations, it cannot be used to geographically identify the boarding location. Past research [139] has used the exact transaction time and the route number to match smart card data to AVL data as a means to identify the location of the transaction. This method cannot be applied to Clipper because 1) the exact boarding time is obfuscated, and 2) the route number is often not recorded. The latter is perhaps more limiting, and due to a combination of the specific technology used, and the institutional challenge of getting the operator to enter the route number at the start of a shift. We suggest that the most reliable means of overcoming this limitation would be a change in technology to a system that automatically records the route and transaction location. In such a system, the transaction time and exact vehicle could be obfuscated to continue to protect privacy. To be cost effective, such a switch would need to occur when new fareboxes are acquired anyway, so there is value in the transit agency being aware of the potential data uses at the time of such a transition.

In spite of this limitation, the research finds that the Clipper data provide value over existing data sources by recording aggregate transfer rates and by providing a consistent form of observation across multiple operators.

Another important issue to consider when working with smart card data is that unless the penetration rate is very high (perhaps higher than 95%), the data reflect potential biases in terms of who chooses to pay with a smart card versus some other medium. When examining the use of Clipper in available onboard transit surveys, the research found several biases, one of which is that that low income and minority travellers were less likely to pay by Clipper. If the data are used for planning studies without accounting for that bias, there is a risk that the needs of these groups will be systematically under-represented.

A correction factor was estimated in an effort to correct this bias. When constrained to use only those variables available in the Clipper data set, the correction factor was found to mitigate, but not eliminate, those biases (they were reduced by about 50%).

The question then is, what can be done about these equity issues? Three sets of approaches might be considered: improve the analysis, increase the use of smart cards among target populations, or mitigate equity issues in the real world. They are not mutually exclusive.

For the analysis, it would be ideal if a better weighting and scheme could be developed that fully accounts for these biases. Including a geographic component to the weight may be beneficial to the weighting scheme, given that low-income and minority travellers are likely concentrated in certain portions of the city. Our analysis is limited because we cannot derive the boarding location for the Clipper card transactions, but if that were known, it may also prove useful in correcting for these other issues. If that is not possible, it may be necessary to bring other, unlinked data for consideration. One can imagine the environmental justice section of a report written to acknowledge this limitation, and also to show overlays of the project location versus low-income and minority populations as recorded by the Census, as an alternative means for identifying the populations that a project might serve.

A second strategy is to make a concerted effort to increase the use of smart cards, especially among target populations. When AC Transit riders who did not use Clipper were asked why, the top reasons were that they preferred cash, they did not want to pay for it or they did not know how to obtain one[144]. This suggests that initiatives to make the cards free, to provide a financial incentive to use the cards or to improve the marketing, outreach and communication of the cards may be effective at increasing their use.

Third, it may be possible to mitigate equity issues in the real world even if the analysis is limited. This can be done by providing free or reduced fares to certain groups (the smart card mechanism enables very detailed targeting to certain populations), or by investing in improvements targeted in certain areas. Such strategies may be part of the "package" necessary for a large project to move forward.

7.1.2 Transit Data Fusion

Chapter 3 described the development and functionality of the core software tool used in this research: a data fusion tool used to measure transit system performance over time. It is implemented for San Francisco, but available under an open source licence and adaptable for use in other regions. The core functionality of the tool is to track transit system performance in four areas: service provided, ridership, level-of-service, reliability and crowding. Performance measures are reported at several levels of resolution, and the software is structured to generate files structured to facilitate model estimation.

There are two distinguishing features of this product.

First, it focuses specifically on monitoring changes over time. In this way, it is useful as a tool for performance based planning, and it provides the data necessary to measure transport project effects in a way that may be useful for informing travel model validation or forecasting.

Second, a key challenge is that the Automated Vehicle Location (AVL) and Automated Passenger Counter (APC) data used in the tool are incomplete, because equipment is installed on only a fraction of buses. To scale these sampled data up to represent the system as a whole, they were merged with General Transit Feed Specification (GTFS) data and weighted based on the ratio of scheduled (bus) trips to recorded trips. This expansion process is similar to expanding a household travel survey to match Census targets of total households by type, and an important strategy for expanding the use of Big Data to locations where only partial or incomplete data are available.

In itself, the use of GTFS data as a means of recording the change in transit schedules is an important enabler of future transport research with a temporal element. Transit routes and schedules are an important determinant of travel patterns, but prior to their recording in a GTFS archive, the details of schedule changes could easily be lost to history. Certainly this is an advance over past norms, as illustrated by one of the first tasks I was given as a travel forecasting intern, in which I was handed a stack of paper bus schedules for New Orleans, and told to code them in MinUTP.

By automating this data processing in a way that can easily be updated as more data become available, this tool enables the analysis of those data for a low incremental cost.

7.1.3 Time Series Model Estimation

In Chapter 4, the outputs of the data tool described in Chapter 3 were used to estimate time series models of transit ridership on the San Francisco Municipal Railway (MUNI) and Bay Area Rapid Transit (BART) systems in the San Francisco Bay Area. Three types of models are considered: Autoregressive Integrated Moving Average (ARIMA) models, regression models and Regression with ARIMA Errors (RegARIMA) models. The ARIMA models themselves do not provide a mechanism to account for known factors that affect ridership. The regression models allow such factors to be considered, but in the case of the BART models, suffer from residual autocorrelation and the risk of spurious regression. Considering both sets of models, the RegARIMA structure performs the best, allowing for a range of predictors to be included, but avoiding problems with residual autocorrelation. It remains challenging to obtain significant parameter estimates for more than a few predictors, probably because a one-dimensional time series allows for a limited number of degrees of freedom, and there can be co-linearity among predictive variables. In spite of this limitation, the models are able to predict ridership as a function of several key variables, including service characteristics, employment changes, fuel cost and an unexplained trend.

7.1.4 Understanding Ridership Trends

Chapter 5 explored divergent ridership trends on two transit systems. Between 2009 and 2013, MUNI ridership decreased by 6.5%, while BART ridership increased by 17.6%. Both systems serve the same region, and employment growth is strong over this period for the markets served by both, so it is surprising that their trajectories are not better aligned.

To understand why, the models estimated in Chapter 4 were applied to break out the factors driving the divergent ridership trends.

The analysis found that the 14.7% of the 17.6% increase in BART ridership could be explained by the terms included in the model. The biggest factor in the increase is employment growth, followed by some modest service increases. There is an additional 2.8% ridership increase beyond what can be explained by the regression terms in the model, which is instead captured as an trend in the ARIMA component of the model.

The analysis found that employment growth had a similar positive effect on MUNI ridership, but that it was offset by two factors. First, was a set of service cuts that served to reduce bus frequency and decrease average runspeed. Second, and more than twice as large, was a downward ridership trend that could not be explained by the terms considered in the model. This trend accounts for a net 11.1% decrease in MUNI ridership, meaning that without it, we would expect MUNI ridership to increase by 4.6%. Several possible explanations were considered to explain this trend, such as a cultural shift towards active transport, the introduction of new shared mobility modes, or the effects of an ageing population. Additional model analysis showed that an increase in the share of the population age 65+ could explain about 60% of this previously unexplained decline. There is weak correlation with other factors, such as the increase in the share of workers commuting by walk, bike, taxi or other modes.

In general, the BART model is more successful at explaining the ridership trends than the MUNI model. One notable difference between the two models is that the BART model is estimated from a much longer time series. It could be that some of the variables tested during estimation and rejected as insignificant would show up as significant when estimated from a longer time series. Doing this may explain more of the trend, but the only way to know for certain would be to try. Alternatively, re-estimating the models as panel data models could potentially improve the significance of additional terms if there is appropriate variation for different stops across time. Finally, it may be that the important factors to consider are difficult to observe or difficult to quantify. For example, we consider whether the reported commute mode share for bike, walk, taxi and other is correlated with the downward trend in MUNI ridership, and find a weak relationship. However, we also know that the use of some of these modes is focused on purposes other than commuting, such as the market for late night social trips via shared mobility modes. More detailed data on the use of and trends in these modes would be beneficial to a better understanding of these issues.

7.1.5 Project-Level Applications

Chapter 6 explored additional applications of the time series models estimated in the previous chapter. The applications are broken into ex-post applications and short term forecasting applications.

In the ex-post applications, the models are applied to measure the change in ridership attributable to each of three separate transit service changes. A key advantage to conducting the evaluation using time series models such as these is that they provide a means for controlling for other factors, beyond the service change, that can be expected to affect ridership. When the results are compared to the results obtained by a simple before-and-after difference which does not control for those factors, the results can be quite different. The results are also compared to the ridership change that would be implied by applying elasticity values found in the literature. The published elasticities tend to have a broad range of values, and while they are a useful point of comparison, the locally estimated values are preferred.

The forecasting applications considered the cases of two transit pilot projects. Following the performance based planning paradigm, each of the pilot projects considered a ridership target that would be set prior to implementation. The analysis showed that the underlying, short-term ridership trends can affect how easy or difficult such a target is to meet, and argues that they should be considered explicitly when setting such targets.

7.2 Future Research Directions

There are several logical extensions to this research, as outlined in this section.

7.2.1 Incorporating Highway Measures

While this work focuses specifically on transit, a logical extension would be to incorporate measures of highway performance.

An initial step in incorporating temporal measures of highway performance at a system level, as opposed to for a specific facility or corridor, is to systematically track the network changes. Often, planning organisations may code new base highway networks in five year increments, but for the resolution desired here, that may be too coarse. It is possible to code networks based on known plans, the management of which can be facilitated by tools such as the San Francisco County Transportation Authority (SFCTA) Network Wrangler [266]. On the transit side, the GTFS provides a convenient mechanism to track network changes. A direct equivalent does not exist on the highway side, although commercial data sets, such as TomTom [267] may fill the need.

Commercial data sets may also be useful both for monitoring speed and reliability [268]. In the Bay Area, loop detector data from the PeMS [94] provides very detailed data from at a large number of freeway locations, but not on surface streets.

Exploratory work has begun using a Global Positioning System (GPS) traces from a fleet of taxi cabs to monitor speed changes in the dense San Francisco network, but that work is not complete at this time.

In terms of observing how demand evolves, traffic counts provide a starting point, although it is more complete if they are available from detectors rather than ad-hoc counts. Origin-destination matrix estimation can be used to translate counts into trip matrices [269], although the problem is not without its challenges [270].

Mobile phone location data also offers a promising path for inferring trip tables [271]. Such an approach was investigated for this study, but data was not available across a sufficiently long time period.

This is an area where there are quite a few pieces in place, and a key research direction is to combine them in an effective manner to learn about the evolution of travel patterns.

7.2.2 Combining with Rolling Household Travel Surveys

A second area where additional data may be advantageous is to take advantage of rolling household travel surveys, such as the UK National Travel Survey (NTS).

Household travel surveys are a much richer travel source in terms of the information available, but their limited sample sizes can make for noisy results when tracking trends. In their analysis of car and train travel trends in Britain, LeVine and Jones [184] managed this issue by grouping multiple years of data, spaced five years apart (1995-1997, 2000-2002 and 2005-2007). A broader question becomes: when considering multiple data sources varying richness and completeness, is it solely up to the analyst to judge how to weight the evidence, or can something more formal be done?

A related issue and opportunity arises in model estimation. Consider the case of estimating mode and destination choice models in Sydney, where multiple waves of data are available. Fox, Patruni and Daly [272] discuss the trade-off between adding more waves of data to increase the sample size, and moving further from the base year for which detailed level-of-service (LOS) data are available. While, on average, the LOS change is probably small from year to year, there may be specific corridors in which a transport project is completed, and the LOS change is important, particularly if it involves a mode becoming newly available for those zone pairs. If annual LOS data were available for a relatively low level of effort, that trade-off would become a non-issue, and more data could be included in the model estimation. This would also fit with the findings of Chapter 5, which found an underlying trend beyond what the model variables could explain. Such a structure in disaggregate estimation could test the inclusion of time-varying variables or constants, and if none are found to be significant, add to the confidence in the stability of the models.

7.2.3 Panel Data Models

As noted in Chapter 4, it was found to be challenging to obtain significant parameter estimates for more than a few explanatory variables using the time series model structure. This may be because a unitary time series does not allow for sufficient independent variation as it relates to those variables. One possible improvement would be to instead use panel data models where there is both cross-sectional and temporal variation permitted in the data. In this instance, an observational unit may be the boardings or alightings at a particular transit stop or station.

In 1990, Kitamura [273] described the virtues of panel analysis in transport planning, which focus largely on the ability to observe changes. His focus was on surveys conducted of the same individuals, although a similar approach could be employed using the type of aggregate data considered here. Tang and Thakuriah [131] and Kerkman et al.[209] provide examples of how such analyses can be used.

7.2.4 Integration into Practice

In addition to furthering the research, it would be valuable to further integrate the types of approaches used in this research into the practice of transport planning. As discussed in Chapter 6, the approaches used here could provide advantages over common practice, and they are not so complicated as to be impractical for use among many planners and modellers. This is particularly important in the public sector, where transport agencies may be sitting on valuable data, but not in a position to take optimal advantage of it. Researchers may have the skills to learn from those data, so effective partnerships may be mutually beneficial. However, there are challenges to establishing such partnerships and working with these data, such as privacy and ownership considerations. These issues are explored further in the next section.

7.3 Lessons Learned

This thesis has explored questions of how to use continuously collected Big Data for the purpose of understanding travel demand trends and measuring transport project impacts. It has done this through specific examples: building a prototype data fusion tool, examining the biases in one specific data set, and applying models to gain insight into specific planning issues. In doing so, we have learned valuable lessons about how to do this sort of work effectively. In this section, the lessons learned from that experience are shared, such that others wishing to continue this sort of work may benefit. These lessons are of particular relevance to transport agencies who may be both sitting on valuable data and facing important planning questions, but still getting their heads around how to best take advantage of those data.

7.3.1 Plan Ahead for Data Archival

The nature of the data considered in this study is that they tend to be continuously collected, providing a record of change in the system. For the planning insights of interest here, a reasonably long record of data is needed, spanning the periods before-and-after transport projects are constructed. However, that record does not begin until someone “hits start”. Until that happens, the data can either be lost for good, or can be very difficult to re-create.

A good example of this situation occurs in the case of real time data feeds. Chapter 1 discussed the `sfdata_collector` software [114], which queries the SFPark Application Program Interface (API) every two minutes to obtain the availability and price of public parking spaces in San Francisco, and write those information to a database. These data can then be analysed to understand trends and patterns. However, the data are not used in the models estimated in Chapter 4 and applied in Chapters 5 and 6, because the record does not begin until after the main 2009 to 2013 analysis period. While a modest effort up front can provide a valuable resource down the road, the time lags are considerable.

French planners appear to have learned this lesson. Ex-post evaluations have been required of major transport projects in France since 1982, but those evaluations were slow to come to fruition due to a number of implementation challenges [274]. One tool introduced to allow such evaluations is the permanent observatory [275]. A permanent observatory is established at the start of a project and records relevant data, including socio-economic information, to allow for the project evaluation later on. An important tension in such observatories balancing the desire to record short lived data, with the complexity and quantity of data recorded.

In a related issue, the data evaluation in Chapter 2 showed how decisions made at the time of contracting or equipment procurement can affect the value of data for years to come. It is important, therefore, to have an awareness of data needs as one consideration among many, such that interventions can be considered at the time when they would be cost-effective.

7.3.2 Big Data is Not a Replacement for Survey Data

A second lesson of this research is that Big Data should be viewed as a complement to travel surveys, not a replacement for travel surveys. For example, Chapter 2 showed the biases that can be found in transit smart card data. It was the availability of onboard transit surveys in the same region that allowed those biases to be detected. If the onboard surveys were not available, we would be in the precarious position of having a data set that we expect to be biased in some way, but which we do not know how.

Even if Big Data were to provide a perfect enumeration of the population, it remains far less rich than travel survey data because it does not record information such as reported trip purpose and mode. This is consistent with the findings of Vij and Shankari [276] who used simulated data sets to determine that larger volumes of data available from GPS-only travel surveys are offset by a loss in data quality.

While this research has demonstrated some ways that Big Data can be used in transport planning, it is important not to sacrifice the data that the industry has come to rely on for years.

7.3.3 Data Access Constraints are a Barrier to Research

Perhaps the biggest risk to the completion of this research successfully was gaining access to the data sets of interest. The most frequently used strategy was to call and ask, but that was contingent on knowing whom to call. In this instance, it was helpful to have worked in the region previously and be able to draw from some pre-existing relationships to find the right points of contact. In spite of this, there were challenges. For example, I was the first person to ask for the Clipper Card data. The technical staff at the Metropolitan Transportation Commission (MTC) were supportive, until the legal team reviewed the initial proposal and added a layer of caution. To their credit, MTC developed a data obfuscation process to accommodate the request [141], but the whole process took about a year of development and negotiation.

To compound the issue, researchers who are working with new and novel data sources are not necessarily incentivised to share those data. There are often good reasons for this, as there may be privacy or licensing restrictions, but there are also competitive incentives to hold data tight.

These access constraints serve as a barrier both to getting a diverse range of people and ideas engaged in this type of research, and to the replicability of the science it produces. This points us towards open data initiatives where they are viable, but it is broadly an issue that we, as a field, must grapple with.

7.3.4 Strategies for Data Privacy

While open data initiatives can be a valuable means for engaging a broader range of analysts, a sometimes competing interest is that of data privacy. Big Data, particularly if it is passively collected data about individuals, poses a very different set of issues than survey data. Central to those issues is the notion of informed consent. When someone responds to a survey, it is a clear and deliberate opt-in on the part of the respondent in a way that does not necessarily occur in the case of passively collected data, even if it is listed in the fine print of a licence agreement. For these reasons, it is of the utmost importance that we act as good stewards of the data with which we are entrusted. Four strategies for protecting data privacy while maintaining the ability to analyse the data to varying degrees are:

Aggregation. If the data released are sufficiently aggregate, it will be impossible to identify individuals.

Obfuscation. This includes strategies such as those used for the Clipper Card data to deliberately repress or "fuzzy" certain fields or information.

Limited distribution. Limiting the distribution to trusted individuals or organisations, perhaps with a data licensing agreement, can serve as a protected measure. In the extreme, there is a big difference between making the data available to an individual researcher versus posting it on the web for anyone to experiment with.

Secure servers. The detailed data can be stored on a secure server, such that researchers can analyse the data on the server, but not remove it. An example is the Transportation Secure Data Center (TSDC) [150].

Each of these strategies involves managing a balance between data protection and the usefulness of those data, so it is valuable for there to be some feedback on how limiting the restrictions end up being.

7.3.5 The Role of Technology Companies

As Big Data becomes trendy, and is increasingly viewed as an asset, commercial interests are looking for ways to monetise that asset. These can be start-ups, such as Urban Engines [259], large technology companies, such as Google or Apple, or firms such as Cubic [277], which is the vendor of the Clipper Card and Oyster Card systems, but is also establishing a data analytics practice to take advantage of the data that comes from those systems. To the degree these firms engage with transport planners as a potential market, they bring a different business model than the engineering firms and software vendors that have traditionally played a large role in transport planning and modelling. It is worth considering how this may shape the profession in the coming years, and how to engage such firms in a way that best serves the public interest.

Consider the case of AirSage [278] as an example. AirSage uses mobile phone location data to generate trip tables and sells those trip tables for the purposes of transport planing and modelling. They have an exclusive arrangement to use data from two of the three major mobile phone carriers in the US, and a proprietary algorithm for processing those data. The result is a very convenient way to get a base year trip table, assuming you are willing to pay the licensing fee.

There are a few questions to ask if you are on the purchasing end.

The first is whether the data are accurate. The black-box nature of the algorithm makes this question difficult to address, so alternative strategies must be considered, such as Huntsinger and Donnelly's work comparing AirSage data to travel model outputs [68]. It is worth considering as well, that the aggregate nature of the data are actually beneficial from a privacy perspective. Any privacy issues occur behind a firewall on AirSage's servers, and the entity purchasing the trip tables does not have direct exposure to those data.

The second question is whether it is worth the price. This must be weighed against the cost of a do-it-yourself solution, if such a solution is possible. Such firms may have access to venture capital, which can exceed the resources available to individual transport planning agencies or university research groups. With a high level of resources, such firms can play an important role in advancing the methods and the state of practice. In some instances, if they hold the data resources, they may be the only entities that can advance the field. This raises questions about what role universities can and should play in such research.

At the same time, it is reasonable to expect that investments in developing such methods would be paid back eventually. This can be mutually beneficial if there is sufficient value to the buyer, but there is also a risk that the price will eventually be based on a "data monopoly", rather than on the analytical value added to those data.

While these are not questions that I claim to answer in this thesis, they are raised here as issues to be faced.

7.4 Implications for Travel Forecasting

This thesis began by considering observed problems with demand forecast accuracy for transport projects, and it finishes with the same. Placing an increased focus on ex-post evaluation of transport projects is one path to improving forecasts, because it provides a basis against which to identify the causes of forecast errors, and it builds a body of reference cases against which forecasts can be compared. Ex-post evaluations have been hampered in the past by the lack of systematic data collection to monitor system changes. Often, transport data are cross-sectional in nature and do not capture changes, or there is a high burden associated with assembling the needed data. Emerging, continuously collected, Big Data were identified as an opportunity to move beyond this barrier.

Using the case of two transit systems in the San Francisco Bay Area, this thesis has demonstrated how such data can be combined to better understand the factors that drive changes in travel demand, and to produce the data streams necessary for ex-post ridership analysis. A key feature is that once such a system is in place, it can be regularly updated as new data become available, and the incremental cost of analysis is low. This opens up a range of additional opportunities to learn from past system changes for the purpose of better predicting the effects of future changes.

There are several additional steps that are necessary to get to the point of improving travel models, notably the evaluation of models against system changes as a means to understand what they get right and what they get wrong. A way forward may be through the Travel Modeling as a Science initiative of the Transportation Research Board (TRB) Committee on Transportation Demand Forecasting. This initiative started from a straw man proposal for improving travel forecasting through a concerted effort that draws from the lessons of the National Oceanic and Atmospheric Administration (NOAA) Hurricane Forecasting Improvement Program [279]. This has evolved into a somewhat more developed proposal [280], an important component of which would be to make available multi-year data for two or more sample cities as a basis against which researchers can judge their model's performance. The goal is to promote models that capture the change, and not just those that calibrate well against base year conditions. This research offers one step towards enabling such a situation.

References

[1] The International Transport Forum. *Spending on Transport Infrastructure 1995-2011: Trends, Policies, Data*. Organisation for Economic Co-operation and Development, May 2013. Available at: `http://www.internationaltransportforum.org/pub/pdf/13SpendingTrends.pdf`.

[2] Pickrell, D. H. *Urban Rail Transit Projects: Forecast Versus Actual Ridership and Cost*. DOT- T -91-04. Urban Mass Transportation Administration, U.S. Department of Transportation, 1990.

[3] Federal Transit Administration. *Before-and-After Studies of New Starts Projects: Report to Congress*. Sept. 2008.

[4] Bain, R. Error and optimism bias in toll road traffic forecasts. *Transportation* 36.5 (2009), pp. 469–482.

[5] Samuel, P. Pie-in-sky traffic and revenue forecasts for Australian tunnel pike hit with class action suits. *Toll Road News* (June 1, 2012).

[6] staff writers. Lawsuits pile up in Oz for toll road forecasts. *Local Transport Today* 653 (Aug. 8, 2014).

[7] Stacey, D. AECOM Says Australia Suit May Hurt. *Wall Street Journal. Business* (Aug. 13, 2014).

[8] Nicolaisen, M. S. and P. A. Driscoll. Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review. *Transport Reviews* 34.4 (2014), pp. 540–557.

[9] Pickrell, D. H. A Desire Named Streetcar Fantasy and Fact in Rail Transit Planning. *Journal of the American Planning Association* 58.2 (1992), pp. 158–176.

[10] Skamris, M. K. and B. Flyvbjerg. Inaccuracy of traffic forecasts and cost estimates on large transport projects. *Transport Policy* 4.3 (1997), pp. 141–146.

[11] Federal Transit Administration. *Light Rail Transit Capital Cost Study Update 2003*. Oct. 2003. Available at: `http://trid.trb.org/view/2003/M/697386`.

[12] Federal Transit Administration and Vanasse Hangen Brustlin. *The Predicted and Actual Impacts of New Starts Projects – 2007: Capital Cost and Ridership*. 2008.

[13] Flyvbjerg, B. Cost Overruns and Demand Shortfalls in Urban Rail and Other Infrastructure. *Transportation Planning and Technology* 30.1 (2007), pp. 9–30.

[14] Flyvbjerg, B. Policy and planning for large-infrastructure projects: problems, causes, cures. *Environment and Planning B: Planning and Design* 34.4 (2007), pp. 578 – 597.

[15] Van Wee, B. Large infrastructure projects: a review of the quality of demand forecasts and cost estimations. *Environment and Planning B: Planning and Design* 34.4 (2007), pp. 611 – 625.

[16] Button, K. J., S. Doh, M. H. Hardy, J. Yuan, and X. Zhou. The Accuracy of Transit System Ridership Forecasts and Capital Cost Estimates. *International Journal of Transport Economics* 37.2 (2010).

[17] Parthasarathi, P. and D. Levinson. Post-construction evaluation of traffic forecast accuracy. *Transport Policy* 17.6 (2010), pp. 428–443.

[18] Welde, M. Demand and operating cost forecasting accuracy for toll road projects. *Transport Policy* 18.5 (2011), pp. 765–771.

[19] Hartgen, D. T. Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling. *Transportation* 40.6 (2013), pp. 1133–1157.

[20] Highways Agency. *Post Opening Project Evaluation of Major Schemes Meta-analysis 2013: Main Report*. 2013, 131p. Available at: `http://assets.highways.gov.uk/our-road-network/pope/major-schemes/pope-meta-2013-final-report.pdf`.

[21] Flyvbjerg, B., N. Bruzelius, and W. Rothengatter. *Megaprojects and risk : an anatomy of ambition*. Cambridge: Cambridge University Press, 2003. 207 pp.

[22] Eliasson, J. and M. Fosgerau. Cost overruns and demand shortfalls – Deception or selection? *Transportation Research Part B: Methodological* 57 (2013), pp. 105–113.

[23] James S. Simpson. *Before-and-After Studies of New Starts Projects: Report to Congress*. Letter. Aug. 10, 2006.

[24] Federal Transit Administration. *Before-and-After Studies of New Starts Projects: Report to Congress*. Sept. 2007.

[25] Ray LaHood. *Before-and-After Studies of New Starts Projects: Report to Congress*. Letter. Dec. 14, 2009.

[26] Ray LaHood. *Before-and-After Studies of New Starts Projects: Report to Congress*. Letter. Nov. 5, 2010.

[27] Federal Transit Administration. *Before-and-After Studies of New Starts Projects: Report to Congress*. Dec. 2011.

[28] Federal Transit Administration. *Before-and-After Studies of New Starts Projects: Report to Congress*. Dec. 2012.

[29] Federal Transit Administration. *Before-and-After Studies of New Starts Projects: Report to Congress*. Dec. 2013.

[30] Flyvbjerg, B. Measuring inaccuracy in travel demand forecasting: methodological considerations regarding ramp up and sampling. *Transportation Research Part A: Policy and Practice* 39.6 (2005), pp. 522–530.

[31] Everett, S., Y. Xiong, K. Sinha C, and J. Fricker D. Ex Post Facto Evaluation of Indiana's Highway Investment Program: Lessons Learned. *Transportation Research Record* 2345 (2013), pp. 24–30.

[32] *Moving Ahead for Progress in the 21st Century Act*. July 6, 2012.

[33] Miller, J. and M. Demetsky. Longitudinal Assessment of Transportation Planning Forecasts. *Journal of Transportation Engineering* 126.2 (2000), pp. 97–106.

[34] Sanko, N., T. Morikawa, and Y. Nagamatsu. Post-project evaluation of travel demand forecasts: Implications from the case of a Japanese railway. *Transport Policy* 27 (2013), pp. 209–218.

[35] Meldrum, M. L. A Brief History of the Randomized Control Trial: From Oranges and Lemons to the Gold Standard. *Hematology/Oncology Clinics of North America* 14.4 (2000), pp. 745–760.

[36] Mass Transportation Commission Staff, McKinsey & Company, Inc., Systems Analysis and Research Corporation, Joseph Napolitan Associates, Inc., and Consultants to the Commission. *Mass Transportation in Massachusetts*. Mass Transportation Commission, Commonwealth of Massachusetts, July 1964. Available at: `https://ia600601.us.archive.org/32/items/masstransportati00mass/masstransportati00mass.pdf`.

[37] Graham-Rowe, E., S. Skippon, B. Gardner, and C. Abraham. Can we reduce car use and, if so, how? A review of available evidence. *Transportation Research Part A: Policy and Practice* 45.5 (2011), pp. 401–418.

[38] Axelrod, R. Advancing the Art of Simulation in the Social Sciences. In: *Handbook of Research on Nature-Inspired Computing for Economics and*

Management. Ed. by Rennard, J.-P. Hersey, PA: Idea Group, 2006, pp. 90–100.

[39] Pratt, R. H. *Traveler response to transportation system changes. Chapter 1-Introduction*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2013. Available at: `http://www.trb.org/Publications/TCRPReport95.aspx`.

[40] Pratt, R. H., N. J. Pedersen, and J. J. Mather. *Traveler response to transportation system changes - A handbook for transportation planners*. Federal Highway Administration, Feb. 1977.

[41] Pratt, R. H. and J. N. Coople. *Traveler response to transportation system changes: Second edition*. Federal Highway Administration, July 1981.

[42] Turnbull, K. F., H. S. Levinson, R. H. Pratt, J. E. Evans, and K. U. Bhatt. *Traveler response to transportation system changes. Chapter 2-HOV Facilities*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2006. Available at: `http://www.trb.org/Publications/TCRPReport95.aspx`.

[43] Turnbull, K. F., R. H. Pratt, J. E. Evans, and H. S. Levinson. *Traveler response to transportation system changes. Chapter 3-Park-and-Ride/Pool*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2004. Available at: `http://www.trb.org/Publications/TCRPReport95.aspx`.

[44] Evans, J. E. and R. H. Pratt. *Traveler response to transportation system changes. Chapter 5-Vanpools and Buspools*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2005. Available at: `http://www.trb.org/Publications/TCRPReport95.aspx`.

[45] Spielberg, F. and R. H. Pratt. *Traveler response to transportation system changes. Chapter 6-Demand Responsive/ADA*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2004. Avail-

able at: http://www.trb.org/Publications/TCRPReport95.aspx.

[46] Evans, J. E. *Traveler response to transportation system changes. Chapter 9-Transit Scheduling and Frequency*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2004. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[47] Pratt, R. H. and J. E. Evans. *Traveler response to transportation system changes. Chapter 10-Bus Routing and Coverage*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2004. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[48] Turnbull, K. F. and R. H. Pratt. *Traveler response to transportation system changes. Chapter 11-Transit Information and Promotion*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2003. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[49] Brian E. McCollom and Richard H. Pratt. *Traveler response to transportation system changes. Chapter 12-Transit Pricing and Fares*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2004. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[50] Vaca, E. and J. R. Kuzmyak. *Traveler response to transportation system changes. Chapter 13-Parking Pricing and Fees*. TCRP 95 Transit Cooperative Research Program. Transportation Research Board, 2005. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[51] Evans, J. E., K. U. Bhatt, and K. F. Turnbull. *Traveler response to transportation system changes. Chapter 14-Road Value Pricing*. TCRP 95 Transit Cooperative Research Program. Transportation Research Board,

2003. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[52] Kuzmyak, J. R., R. H. Pratt, G. B. Douglas, and F. Spielberg. *Traveler response to transportation system changes. Chapter 15-Land Use and Site Design*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2003. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[53] Pratt, R. H., J. E. Evans, H. S. Levinson, S. M. Turner, C. Y. (Jeng, and D. Nabors. *Traveler response to transportation system changes. Chapter 16-Pedestrian and Bicycle Facilities*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2012. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[54] Evans, J. E., R. H. Pratt, A. Stryker, and J. R. Kuzmyak. *Traveler response to transportation system changes. Chapter 17-Transit Oriented Development*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2007. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[55] Kuzmyak, J. R., R. Weinberger, R. H. Pratt, and H. S. Levinson. *Traveler response to transportation system changes. Chapter 18-Parking Management and Supply*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2003. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[56] Kuzmyak, J. R., J. E. Evans, and R. H. Pratt. *Traveler response to transportation system changes. Chapter 19-Employer and Institutional TDM Strategies*. TCRP Report 95 Transit Cooperative Research Program. Transportation Research Board, 2010. Available at: http://www.trb.org/Publications/TCRPReport95.aspx.

[57] Olsson, N. O. E., H. P. Krane, A. Rolstadås, and M. Veiseth. Influence of reference points in ex post evaluations of rail infrastructure projects. *Transport Policy* 17.4 (2010), pp. 251–258.

[58] Shuldiner, A. T. and P. W. Shuldiner. The measure of all things: reflections on changing conceptions of the individual in travel demand modeling. *Transportation* 40.6 (2013), pp. 1117–1131.

[59] Wolf, J., W. Bachman, M. Oliveira, J. Auld, A. Mohammadian, and P. Vovsha. *Applying GPS Data to Understand Travel Behavior. Volume I.* NCHRP Report 775 National Cooperative Highway Research Program. Transportation Research Board, 2014.

[60] Wolf, J., W. Bachman, M. Oliveira, J. Auld, A. Mohammadian, P. Vovsha, and J. Zmud. *Applying GPS Data to Understand Travel Behavior. Volume II.* NCHRP Report 775 National Cooperative Highway Research Program. Transportation Research Board, 2014.

[61] Liu, H. X. and W. Ma. A virtual vehicle probe model for time-dependent travel time estimation on signalized arterials. *Transportation Research Part C: Emerging Technologies* 17.1 (2009), pp. 11–26.

[62] Pascual, O. S. Statistical Complex Analysis of Taxi Mobility in San Francisco. Master's Thesis. Universitat de Barcelona, 2011.

[63] Castro, P. S., D. Zhang, and S. Li. Urban Traffic Modelling and Prediction Using Large Scale Taxi GPS Traces. In: *Pervasive Computing*. Ed. by Kay, J., P. Lukowicz, H. Tokuda, P. Olivier, and A. Krüger. Lecture Notes in Computer Science 7319. Springer Berlin Heidelberg, Jan. 1, 2012, pp. 57–72.

[64] Zhan, X., S. Hasan, S. V. Ukkusuri, and C. Kamga. Urban link travel time estimation using large-scale taxi data with partial information. *Transportation Research Part C: Emerging Technologies* 33 (2013), pp. 37–49.

[65] Jiang, S., J. Ferreira , Jr., and M. C. Gonzalez. Discovering urban spatial-temporal structure from human activity patterns. In: *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*. UrbComp '12. New York, NY, USA: ACM, 2012, pp. 95–102.

[66] Wang, P., T. Hunter, A. M. Bayen, K. Schechtner, and M. C. Gonzalez. Understanding Road Usage Patterns in Urban Areas. *Scientific Reports* 2 (2012).

[67] Aguiléra, V., S. Allio, V. Benezech, F. Combes, and C. Milion. Using cell phone data to measure quality of service and passenger flows of Paris transit system. *Transportation Research Part C: Emerging Technologies*. Special Issue with Selected Papers from Transport Research Arena 43, Part 2 (2014), pp. 198–211.

[68] Huntsinger, L. F. and R. Donnelly. Reconciliation of Regional Travel Model and Passive Device Tracking Data. In: Transportation Research Board Annual Meeting. Washington, D.C., 2014.

[69] Wang, T., C. Chen, and J. Ma. Mobile Phone Data as an Alternative Data Source for Travel Behavior Studies. In: Transportation Research Board Annual Meeting. Washington, D.C., 2014.

[70] Çolak, S., L. P. Alexander, B. G. Alvim, S. R. Mehndiretta, and M. C. Gonzalez. Analyzing Cell Phone Location Data for Urban Travel: Current Methods, Limitations and Opportunities. In: Transportation Research Board Annual Meeting. Washington, D.C., 2015.

[71] Transportation Research Board. *NCHRP 08-95*. May 30, 2014. [ONLINE]: `http://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=3637` (visited on 10/30/2014).

[72] Haghani, A. and M. Hamedi. Application of Bluetooth Technology in Traffic Detection, Surveillance, and Traffic Management. *Journal of Intelligent Transportation Systems* 17.2 (2013), pp. 107–109.

[73] Kim, J., F. Kurauchi, N. Uno, T. Hagihara, and T. Daito. Using Electronic Toll Collection Data to Understand Traffic Demand. *Journal of Intelligent Transportation Systems* 18.2 (2013), pp. 190–203.

[74] Bertini, R. L., S. Hansen, A. Byrd, and T. Yin. Experience implementing a user service for archived intelligent transportation systems data. *Transportation Research Record* 1917.1 (2005), pp. 90–99.

[75] Yuan, Y., H. Van Lint, F. Van Wageningen-Kessels, and S. Hoogendoorn. Network-Wide Traffic State Estimation Using Loop Detector and Floating Car Data. *Journal of Intelligent Transportation Systems* 18.1 (2013), pp. 41–50.

[76] Himpe, W. W. E., C. M. J. Tampère, and B. Moelans. A Parsimonious Method for Offline Freeway Travel Time Estimation From Sectional Speed Detectors. *Journal of Intelligent Transportation Systems* 18.1 (May 13, 2013), pp. 67–80.

[77] Berkow, M., J. Chee, R. L. Bertini, and C. Monsere. Transit performance measurement and arterial travel time estimation using archived AVL data. In: ITE District 6 Annual Meeting. Portland, Oregon, July 2007.

[78] Tétreault, P. R. and A. M. El-Geneidy. Estimating bus run times for new limited-stop service using archived AVL and APC data. *Transportation Research Part A: Policy and Practice* 44.6 (2010), pp. 390–402.

[79] Byon, Y.-J., C. E. Cortés, M. C, F. Javier, M. Munizaga, and M. Zuniga. Transit Performance Monitoring and Analysis with Massive GPS Bus Probes of Transantiago in Santiago, Chile: Emphasis on Development of Indices for Bunching and Schedule Adherence. In: Transportation Research Board Annual Meeting. Washington, D.C., 2011.

[80] El-Geneidy, A. M., J. Horning, and K. J. Krizek. Analyzing transit service reliability using detailed data from automatic vehicular locator systems. *Journal of Advanced Transportation* 45.1 (2011), pp. 66–79.

[81] Furth, P. G., B. Hemily, T. H. Muller, and J. G. Strathman. *Using Archived AVL-APC Data to Improve Transit Performance and Management*. TCRP Report 113 Transit Cooperative Research Program. Transportation Research Board, 2006. Available at: `http://www.nap.edu/catalog.php?record_id=13907`.

[82] Cevallos, F. and X. Wang. Adams Data Archiving and Mining System for Transit Service Improvements. *Transportation Research Record* 2063.-1 (2008), pp. 43–51.

[83] Chu, X. *A Guidebook for Using Automatic Passenger Counter Data for National Transit Database (NTD) Reporting*. National Center for Transit Research (NCTR) University of South Florida, Dec. 2010. Available at: `http://www.dot.state.fl.us/research-center/Completed_Proj/Summary_PTO/FDOT_BDK85_977-04_rpt.pdf`.

[84] Liao, C.-F. *Data Driven Support Tool for Transit Data Analysis, Scheduling and Planning*. Intelligent Transportation Systems Institute, Center for Transportation Studies, University of Minnesota, 2011. Available at: `http://conservancy.umn.edu/bitstream/11299/149402/1/CTS11-15.pdf`.

[85] Barry, J., R. Freimer, and H. Slavin. Use of Entry-Only Automatic Fare Collection Data to Estimate Linked Transit Trips in New York City. *Transportation Research Record* 2112.-1 (2009), pp. 53–61.

[86] Reddy, A., A. Lu, S. Kumar, V. Bashmakov, and S. Rudenko. Entry-Only Automated Fare-Collection System Data Used to Infer Ridership, Rider Destinations, Unlinked Trips, and Passenger Miles. *Transportation Research Record* 2110.-1 (2009), pp. 128–136.

[87] Pelletier, M.-P., M. Trépanier, and C. Morency. Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies* 19.4 (2011), pp. 557–568.

[88] Gordon, J., H. Koutsopoulos, N. Wilson, and J. Attanucci. Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data. *Transportation Research Record* 2343.-1 (2013), pp. 17–24.

[89] Wilhelm, J., J. Wolf, E. Kang, and D. Taylor. The Cleveland GPS Household Travel Survey: Survey Design, Imputation of Trip Characteristics, and Secondary Uses of the Data. In: Transportation Research Board Annual Meeting. Washington, D.C., 2014.

[90] Zmud, J., M. Lee-Gosselin, J. Carrasco, and M. A. Munizaga. *Transport Survey Methods: Best Practice for Decision Making*. Emerald Group Publishing Limited, 2013. 798 pp.

[91] Smith, T. W. Survey-Research Paradigms Old and New. *International Journal of Public Opinion Research* 25.2 (2013), pp. 218–229.

[92] *General Transit Feed Specification Reference - Transit — Google Developers*. [ONLINE]: `https://developers.google.com/transit/gtfs/reference` (visited on 07/15/2014).

[93] *GTFS Data Exchange - San Francisco Municipal Transportation Agency*. [ONLINE]: `http://www.gtfs-data-exchange.com/agency/san-francisco-municipal-transportation-agency/` (visited on 07/15/2014).

[94] Caltrans. *Caltrans PeMS*. [ONLINE]: `http://pems.dot.ca.gov/` (visited on 07/15/2014).

[95] The Exploratorium, Yellow Cab, and Stamen Design. *Cabspotting*. [ONLINE]: `http://cabspotting.org/` (visited on 11/04/2014).

[96] Tischler, D., L. Zorn, and E. Sall. Creating CountDracula: An Open Source Counts Management Tool. In: Transportation Research Board Annual Meeting. Washington, D.C., 2014.

[97] Schrank, D., B. Eisele, T. Lomax, and J. Bak. *2015 Urban Mobility Scorecard*. Texas A&M Transportation Institute and INRIX, Inc., 2015. Available at: `http://mobility.tamu.edu/ums/report/`.

[98] *TomTom Traffic Index*. 2016. [ONLINE]: `https://www.tomtom.com/en_gb/trafficindex/` (visited on 07/02/2016).

[99] Cambridge Systematics, Vanasse Hangen Brustlin, Gallop Corporation, Chandra R. Bhat, Shapiro Transportation Consulting, and Martin/Alexiou/Bryson. *Travel demand forecasting: parameters and techniques*. NCHRP Report 716. Transportation Research Board, 2012.

[100] Horowitz, A. J. *Statewide travel forecasting models*. NCHRP Synthesis 358 National Cooperative Highway Research Program. Transportation Research Board, 2006.

[101] US Census Bureau Center for Economic Studies. *Longitudinal Employer-Household Dynamics*. [ONLINE]: `http://lehd.ces.census.gov/data/` (visited on 11/04/2014).

[102] U.S. Census Bureau. *American Community Survey*. [ONLINE]: `http://www.census.gov/acs/www/` (visited on 11/05/2014).

[103] Energy Information Administration. *Gasoline and Diesel Fuel Update*. [ONLINE]: `http://www.eia.gov/petroleum/gasdiesel/` (visited on 11/05/2014).

[104] Internal Revenue Service. *Standard Mileage Rates*. [ONLINE]: `http://www.irs.gov/Tax-Professionals/Standard-Mileage-Rates` (visited on 11/05/2014).

[105] Bay Area Toll Authority. *Toll Schedule for State-Owned Toll Bridges*. [ONLINE]: `http://bata.mtc.ca.gov/tolls/schedule.htm` (visited on 11/05/2014).

[106] Oracle Consulting. *SFpark service API reference*. Dec. 17, 2013. [ONLINE]: `http://sfpark.org/resources/sfpark-service-api-reference-updated/` (visited on 11/03/2014).

[107] Bureau of Labor Statistics. *Consumer Price Index (CPI)*. [ONLINE]: http://www.bls.gov/cpi/ (visited on 11/05/2014).

[108] NuStats, GeoStats, Franklin Hill Group, and Mark Bradley Research & Consulting. *2010-2012 California Household Travel Survey: Final Report*. California Department of Transportation, June 2013.

[109] Jonnalagadda, N., J. Freedman, W. A. Davidson, and J. D. Hunt. Development of microsimulation activity-based model for San Francisco: destination and mode choice models. *Transportation Research Record: Journal of the Transportation Research Board* 1777.1 (2001), pp. 25–35.

[110] Erhardt, G., B. Charlton, J. Freedman, J. Castiglione, and M. Bradley. Enhancement and application of an activity-based travel model for congestion pricing. In: Transportation Research Board Innovations in Travel Modeling Conference. Portland, Oregon, 2008.

[111] Zorn, L., E. Sall, and D. Wu. Incorporating crowding into the San Francisco activity-based travel model. *Transportation* 39.4 (2012), pp. 755–771.

[112] Erhardt, G. D., E. Sall, L. Zorn, D. Tischler, R. Alsup, and N. Nassir. Development and Application of a Dynamic Traffic Assignment Model for San Francisco. In: European Transport Conference. Frankfurt, Germany, 2013.

[113] Parsons Brinckerhoff and San Francisco County Transportation Authority. *San Francisco Dynamic Traffic Assignment Project, "DTA Anyway": Final Methodology Report*. Nov. 28, 2012. Available at: https://code.google.com/p/dta/.

[114] Erhardt, G. D. *sfdata_collector*. 2014. [ONLINE]: https://github.com/UCL/sfdata_collector (visited on 11/03/2014).

[115] Erhardt, G. D. *sfdata_wrangler*. 2014. [ONLINE]: https://github.com/UCL/sfdata_wrangler (visited on 07/15/2014).

[116] The HDF Group. *HDF Group - HDF5*. [ONLINE]: http://www.hdfgroup.org/HDF5/ (visited on 11/03/2014).

[117] Lock, O. C. The use of automatically collected transport data in the spatiotemporal analysis and visualisation of policy change: a case study of the San Francisco Municipal Railway. MRes Thesis. University College London (UCL), 2014.

[118] *Python Data Analysis Library — pandas: Python Data Analysis Library*. [ONLINE]: `http://pandas.pydata.org/` (visited on 11/03/2014).

[119] Frank Haight. Principles of evaluation. In: *Project and policy evalution in transport*. Ed. by Liana Giorgi and A. D. Pearman. Aldershot: Ashgate, 2002.

[120] Institute of Transportation Engineers Transportation Safety Council. *Before-and-after study technical brief*. May 2009. Available at: `http://www.cite7.org/documents/Before_After%20Study_Published.pdf`.

[121] Coglianese, C. *Measuring Regulatory Performance*. Expert Paper No. 1. Organisation for Economic Co-operation and Development, Aug. 2012. Available at: `http://www.oecd.org/regreform/regulatory-policy/1_coglianese%20web.pdf`.

[122] Tay, R., A. Azad, S. C. Wirasinghe, and S. Hansen. Analysis of the influence of urban rail stations on neighbourhood crime. *International Journal of Urban Sciences* 17.2 (2013), pp. 281–289.

[123] De Grange, L., R. Troncoso, and F. González. An empirical evaluation of the impact of three urban transportation policies on transit use. *Transport Policy* 22 (2012), pp. 11–19.

[124] Ogilvie, D., C. E. Foster, H. Rothnie, N. Cavill, V. Hamilton, C. F. Fitzsimons, and N. Mutrie. Interventions to promote walking: systematic review. *BMJ* 334.7605 (2007), p. 1204.

[125] Economic Development Research Group. *Transportation Impacts of the Massachusetts Turnpike Authority and the Central Artery/Third Harbor*

Tunnel Project. Massachusetts Turnpike Authority, Feb. 2006. Available at: http://www.edrgroup.com/pdf/mta-economic-v1.pdf.

[126] Economic Development Research Group. *Real Estate Impacts of the Massachusetts Turnpike Authority and the Central Artery/Third Harbor Tunnel Project*. Massachusetts Turnpike Authority, Feb. 2006. Available at: http://www.edrgroup.com/pdf/mta-economic-v2.pdf.

[127] Fitzroy, S., G. Weisbrod, and N. Stein. Ex-post Assessment of Transport Investments and Policy Interventions. In: Roundtable on Ex-post Assessment of Transport Investments and Policy Interventions. Paris: The International Transport Forum at the OECD, Sept. 15–16, 2014.

[128] Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. *Time series analysis : forecasting and control*. Fourth edition. Hoboken, NJ: JWiley & Sons, 2008.

[129] Chen, C., D. Varley, and J. Chen. What Affects Transit Ridership? A Dynamic Analysis involving Multiple Factors, Lags and Asymmetric Behaviour. *Urban Studies* 48.9 (2011), pp. 1893–1908.

[130] Diggle, P. *Analysis of longitudinal data*. 2nd ed. Oxford: Oxford University Press, 2002. 379 pp.

[131] Tang, L. and P. Thakuriah. Ridership effects of real-time bus information system: A case study in the City of Chicago. *Transportation Research Part C: Emerging Technologies* 22 (2012), pp. 146–161.

[132] Frazier, C. and K. Kockelman. Spatial Econometric Models for Panel Data: Incorporating Spatial and Temporal Data. *Transportation Research Record* 1902 (2005), pp. 80–90.

[133] Cheng, T., J. Haworth, and J. Wang. Spatio-temporal autocorrelation of road network data. *Journal of Geographical Systems* 14.4 (2012), pp. 389–413.

[134] Graham, D. J. Causal Influence for Ex-post Evaluation of Transport Interventions. In: Roundtable on Ex-post Assessment of Transport Investments and Policy Interventions. Paris: International Transport Forum at the Organisation for Economic Co-operation and Development, Sept. 15–16, 2014.

[135] Li, H., D. J. Graham, and A. Majumdar. The effects of congestion charging on road traffic casualties: A causal analysis using difference-in-difference estimation. *Accident Analysis & Prevention* 49 (2012), pp. 366–377.

[136] Medina, S. and A. Erath. Estimating Dynamic Workplace Capacities by Means of Public Transport Smart Card Data and Household Travel Survey in Singapore. *Transportation Research Record* 2344.-1 (2013), pp. 20–30.

[137] Zhong, C., S. M. Arisona, X. Huang, M. Batty, and G. Schmitt. Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science* (2014).

[138] Zhong, C., E. Manley, S. Müller Arisona, M. Batty, and G. Schmitt. Measuring variability of mobility patterns from multiday smart-card data. *Journal of Computational Science* 9 (2015), pp. 125–130.

[139] Munizaga, M. A. and C. Palma. Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies* 24 (2012), pp. 9–18.

[140] Wang, Z.-J., X.-H. Li, and F. Chen. Impact evaluation of a mass transit fare change on demand and revenue utilizing smart card data. *Transportation Research Part A: Policy and Practice* 77 (2015), pp. 213–224.

[141] Ory, D. Lawyers, Big Data, (More Lawyers), and a Potential Validation Source: Obtaining Smart Card and Toll Tag Transaction Data. In: Transportation Research Board Annual Meeting. Washington, D.C, 2015.

[142] Metropolitan Transportation Commission. *Statistical Summary of Bay Area Transit Operators: Fiscal Years 2009-10 Through 2013-14*. July 2015. Available at: `http://mtc.ca.gov/sites/default/files/StatSumm_2014.pdf`.

[143] BART. *Monthly Ridership Reports*. [ONLINE]: `http://www.bart.gov/about/reports/ridership` (visited on 08/01/2015).

[144] Redhill Group, Inc. *AC Transit 2012 Passenger Survey: Draft Findings Report*. Metropolitan Transportation Commission, Feb. 6, 2013.

[145] Redhill Group, Inc. *SamTrans 2013 Passenger Study: Draft Survey Findings Report*. Metropolitan Transportation Commission, Aug. 12, 2013.

[146] Redhill Group, Inc. *Golden Gate Ferry 2013 Passenger Study: Draft Survey Findings Report*. Metropolitan Transportation Commission, Dec. 23, 2013.

[147] Redhill Group, Inc. *Golden Gate Transit 2013 Passenger Study: Draft Survey Findings Report*. Metropolitan Transportation Commission, Dec. 23, 2013.

[148] Redhill Group, Inc. *San Francisco Bay Ferry 2013 Passenger Study: Draft Survey Findings Report*. Metropolitan Transportation Commission, Jan. 21, 2014.

[149] *Caltrain 2014 Passenger Study: Draft Survey Findings Report*. Prepared for Caltrain and MTC, 2014.

[150] Gonder, J. D., E. Burton, and E. Murakami. Archiving Data from New Survey Technologies: Lessons Learned on Enabling Research with High-Precision Data While Preserving Participant Privacy. In: 10th International Conference on Transport Survey Methods. Leura, Australia, Nov. 16–21, 2014.

[151] Turnbull, K. F. Conference Summary. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Ed. by Turnbull, K. F. Irvine, California: Transportation Research Board, 2013.

[152] Lomax, T., R. S. Blankenhorn, and R. Watanabe. Clash of Priorities. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Irvine, California: Transportation Research Board, 2013.

[153] Price, T. J., D. Miller, C. Fulginiti, and S. Terabe. Performance-Based Decision Making: The Buck Starts Here. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Irvine, California: Transportation Research Board, 2013.

[154] Benson, J. R., R. Perrin, and S. M. Pickrell. Measuring Transportation System and Mode Performance. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Irvine, California: Transportation Research Board, 2013.

[155] Winick, R. M., W. Bachman, Y. Sekimoto, and P. S. Hu. Transforming Experiences: From Data to Measures, Measures to Information, and Information to Decisions with Data Fusion and Visualization. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Irvine, California: Transportation Research Board, 2013.

[156] Zmud, J., A. Brush, and M. D. Choudhury. Digital Breadcrumbs: Mobility Data Capture with Social Media. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Irvine, California: Transportation Research Board, 2013.

[157] Pack, M. Asking the Right Questions: Timely Advice for Emerging Tools, Better Data, and Approaches for Systems Performance Measures. In: *Performance measurement of transportation systems: summary of the Fourth International Conference*. Irvine, California: Transportation Research Board, 2013.

[158] Grant, M., J. D'Ignazio, A. Bond, and A. McKeeman. *Performance-Based Planning and Programming Guidebook*. FHWA-HEP-13-041. Federal Highway Administration, Sept. 2013. Available at: `http://www.fhwa.dot.gov/planning/performance_based_planning/pbpp_guidebook/pbppguidebook.pdf`.

[159] Furth, P. G. *Data analysis for bus planning and monitoring*. TCRP Synthsis 34. Transportation Research Board, 2000. Available at: `http://onlinepubs.trb.org/onlinepubs/tcrp/tsyn34.pdf`.

[160] Kittelson & Associates, Urbitran, LKC Consulting Services, MORPACE International, Queensland University of Technology, and Yuko Nakanishi. *A guidebook for developing a transit performance-measurement system*. TCRP Report 88 Transit Cooperative Research Program. Transportation Research Board, 2003. Available at: `http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_report_88/Guidebook.pdf`.

[161] Liao, C.-F. and H. Liu. Development of Data-Processing Framework for Transit Performance Analysis. *Transportation Research Record: Journal of the Transportation Research Board* 2143.-1 (2010), pp. 34–43.

[162] Feng, W. and M. Figliozzi. Using Archived AVL/APC Bus Data to Identify Spatial-Temporal Causes of Bus Bunching. In: Transportation Research Board Annual Meeting. Washington, D.C., 2011.

[163] Berkow, M., A. El-Geneidy, R. Bertini, and D. Crout. Beyond Generating Transit Performance Measures: Visualization and Statistical Analysis with Historic Data. *Transportation Research Record* 2111.-1 (2009), pp. 158–168.

[164] Mesbah, M., G. Currie, C. Lennon, and T. Northcott. Spatial and temporal visualization of transit operations performance data at a network level. *Journal of Transport Geography*. Special Section on Accessibility and Socio-Economic Activities: Methodological and Empirical Aspects 25 (2012), pp. 15–26.

[165] Wang, J., Y. Li, J. Liu, K. He, and P. Wang. Vulnerability Analysis and Passenger Source Prediction in Urban Rail Transit Networks. *PLoS ONE* 8.11 (2013), e80178.

[166] Chen, W.-Y. and Z.-Y. Chen. A Simulation Model for Transit Service Unreliability Prevention Based on AVL-APC Data. In: International Conference on Measuring Technology and Mechatronics Automation. Vol. 2. Apr. 2009, pp. 184–188.

[167] Harris, K. D. *Making Your Privacy Practices Public: Recommendations on Developing a Meaningful Privacy Policy*. California Department of Justice, May 2014. Available at: `http://www.nowlandlaw.com/docs/making_your_privacy_practices_public.pdf`.

[168] Batty, M. Big data, smart cities and city planning. *Dialogues in Human Geography* 3.3 (2013), pp. 274–279.

[169] Kittelson & Associates, Parsons Brinckerhoff, KFH Group, Inc., Texas A&M Transportation Institute, and Arup. *Transit capacity and quality of service manual*. TCRP Report 165 Transit Cooperative Research Program. Transportation Research Board, 2013. Available at: `http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_165fm.pdf`.

[170] Taylor, B. D. and C. N. Fink. *The Factors Influencing Transit Ridership: A Review and Analysis of the Ridership Literature*. Working Paper. UCLA Department of Urban Planning, 2003. Available at: `http://www.uctc.net/research/papers/681.pdf`.

[171] Hendrickson, C. A note on trends in transit commuting in the United States relating to employment in the central business district. *Transportation Research Part A: General* 20.1 (1986), pp. 33–37.

[172] McLeod Jr, M. S., K. J. Flannelly, L. Flannelly, and R. W. Behnke. Multivariate Time-Series Model Of Transit Ridership Based On Historical, Aggregate Data: The Past, Present And Future Of Honolulu. *Transportation Research Record* 1297 (1991).

[173] Gomez-Ibanez, J. A. Big-city transit, ridership, deficits, and politics. *American Planning Association. Journal of the American Planning Association* 62.1 (1996), p. 30.

[174] Kain, J. F. and Z. Liu. Secrets of success: assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. *Transportation Research Part A: Policy and Practice* 33.7 (1999), pp. 601–624.

[175] Brown, J. R. and D. Neog. Urban structure and transit ridership: A reexamination of the relationship in the United States. In: Transportation Research Board Annual Meeting. Washington, D.C., 2007.

[176] Taylor, B. D., D. Miller, H. Iseki, and C. Fink. Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice* 43.1 (2009), pp. 60–77.

[177] Haire, A. and R. Machemehl. Regional and Modal Variability in Effects of Gasoline Prices on U.S. Transit Ridership. *Transportation Research Record* 2144 (2010), pp. 20–27.

[178] Lane, B. W. A time-series analysis of gasoline prices and public transportation in US metropolitan areas. *Journal of Transport Geography*. Special Section on Rail Transit Systems and High Speed Rail 22 (May 2012), pp. 221–235.

[179] Nowak, W. P. and I. Savage. The cross elasticity between gasoline prices and transit use: Evidence from Chicago. *Transport Policy* 29 (2013), pp. 38–45.

[180] Yanmaz-Tuzel, O. and K. Ozbay. Impacts of Gasoline Prices on New Jersey Transit Ridership. *Transportation Research Record* 2144 (2010), pp. 52–61.

[181] Anderson, M. L. *Subways, Strikes, and Slowdowns: The Impacts of Public Transit on Traffic Congestion*. 18757 Working Paper. National Bureau of Economic Research, Feb. 2013. Available at: `http://www.nber.org/papers/w18757`.

[182] Brakewood, C., G. S. Macfarlane, and K. Watkins. The impact of real-time information on bus ridership in New York City. *Transportation Research Part C: Emerging Technologies* 53 (Apr. 2015).

[183] Metz, D. Demographic determinants of daily travel demand. *Transport Policy* 21 (2012), pp. 20–25.

[184] Le Vine, S. and P. Jones. *Making sense of car and train travel trends in Britain*. RAC Foundation, Dec. 2012. Available at: `http://nudgeadvisory.com/assets/uploaded/downloads/On_The_Move_Making_Sense_of_Car_and_Train_Trends_in_Britain.pdf`.

[185] Gordon, R. and M. Cabanatuan. Muni looks at some of deepest service cuts ever. *San Francisco Chronicle* (Jan. 29, 2010).

[186] San Francisco Municipal Transportation Agency. *All-Door Boarding Evaluation: Final Report*. Dec. 2014. Available at: `https://www.sfmta.com/sites/default/files/agendaitems/2014/12-2-14%20Item%2014%20All%20Door%20Boarding%20Report.pdf`.

[187] Metropolitan Transportation Commission, San Francisco County Transportation Authority, and Puget Sound Regional Council. *Moving Fast-Trips from Research to Practice...a three-agency experience*. [ONLINE]: `http://fast-trips.mtc.ca.gov/` (visited on 04/14/2016).

[188] Khani, A., E. Sall, and M. Hickman. Integration of the FAST-TrIPs Person-Based Dynamic Transit Assignment Model, the SF-CHAMP Regional, Activity-Based Travel Demand Model, and San Francisco's Citywide Dynamic Traffic Assignment Model. In: Transportation Research Board Annual Meeting. Washington, DC, 2013.

[189] Metropolitan Transportation Commission, San Francisco County Transportation Authority, and Puget Sound Regional Council. *SHRP2 Implementation Assistance Program Round 4 Application Form*. June 27, 2014.

[190] *Quarterly Census of Employment and Wages*. [ONLINE]: http://www.bls.gov/cew/ (visited on 04/14/2016).

[191] Gumbel, A. San Francisco's guerrilla protest at Google buses swells into revolt. *The Guardian* (Jan. 25, 2014).

[192] Colliers International. *North America CBD Parking Rate Survey Highlights*. 2007.

[193] Colliers International. *CBD Parking Rate Survey*. 2008.

[194] Colliers International. *Parking Rates: Global CBD Parking Rate Survey*. 2009.

[195] Colliers International. *North America Central Business District Parking Rate Survey*. 2010.

[196] Colliers International. *Global Central Business District Parking Rate Survey*. 2011.

[197] Colliers International. *North America Central Business District 2012 Parking Rate Survey*. 2012.

[198] Hyndman, R. J. and G. Athanasopoulos. *Forecasting principles and practice*. OTexts, Oct. 17, 2013. 292 pp.

[199] Ljung, G. M. and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika* 65.2 (Jan. 8, 1978), pp. 297–303.

[200] Rob J. Hyndman and Yeasmin Khandakar. Automatic Time Series Forecasting: The forecast Package for R. *Journal of Statistical Software* 27.3 (2008).

[201] Gardner, G., A. C. Harvey, and G. D. A. Phillips. Algorithm AS 154: An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 29.3 (1980), pp. 311–322. JSTOR: 2346910.

[202] Washington, S. P., M. G. Karlaftis, and F. L. Mannering. *Statistical and Econometric Methods for Transportation Data Analysis, Second Edition*. CRC Press, Dec. 2, 2010. 546 pp.

[203] Jones, C., K. Williams, K. V. Brown, and J. Garofoli. BART strike has transit, commuters scrambling. *San Francisco Chronicle* (July 1, 2013).

[204] Wildermuth, J., M. Cabanatuan, and J. Tucker. BART workers go on strike. *San Francisco Chronicle* (Oct. 18, 2013).

[205] Chris Rauber. Has BART reached its breaking point? *San Francisco Business Times* (July 10, 2015).

[206] Cervero, R., J. Murakami, and M. Miller. Direct Ridership Model of Bus Rapid Transit in Los Angeles County, California. *Transportation Research Record* 2145 (2010), pp. 1–7.

[207] Gutiérrez, J., O. D. Cardozo, and J. C. García-Palomares. Transit ridership forecasting at station level: an approach based on distance-decay weighted regression. *Journal of Transport Geography*. Special section on Alternative Travel futures 19.6 (2011), pp. 1081–1092.

[208] Liu, C., S. Erdogan, T. Ma, and F. W. Ducca. How to Increase Rail Ridership in Maryland? Direct Ridership Models (DRM) for Policy Guidance. In: Transportation Research Board Annual Meeting. Washington, D.C., 2014.

[209] Kerkman, K., K. Martens, and H. Meurs. Factors Influencing Stop Level Transit Ridership in the Arnhem Nijmegen City Region. In: Transportation Research Board Annual Meeting. Washington, D.C., 2015.

[210] Stephen Levy. *Bay Area Job Growth to 2040: Projections and Analysis*. Association of Bay Area Governments, Feb. 2012. Available at: `http://planbayarea.org/pdf/3-9-12/CCSCE_Bay_Area_Job_Growth_to_2040.pdf`.

[211] Parsons Brinckerhoff. *BART Seismic Impact Analysis: Traffic Forecasting Report*. Bay Area Rapid Transit, 2013.

[212] Erhardt, G., D. Ory, A. Sarvepalli, J. Freedman, J. Hood, and B. Stabler. MTC's Travel Model One: Applications of an Activity-Based Model in its First Year. In: *5th Transportation Research Board Innovations in Travel Modeling Conference,* 5th Transportation Research Board Innovations in Travel Modeling Conference, Tampa, Florida, 2012.

[213] Rosenberger, T., C. Nardi, and K. Liwag. What is Generating Transit Ridership Increase in US and Canadian Cities? In: *European Transport Conference*. European Transport Conference. 2015.

[214] Primus, J. *Reversing Muni's Downward Spiral*. SPUR, May 18, 2005. Available at: `http://www.spur.org/publications/spur-report/2005-05-18/reversing-munis-downward-spiral`.

[215] Metz, D. Saturation of Demand for Daily Travel. *Transport Reviews* 30.5 (Sept. 1, 2010), pp. 659–674.

[216] Goodwin, P. *Peak Travel, Peak Car and the Future of Mobility: Evidence, Unresolved Issues, Policy Implications, and a Research Agenda*. Organisation for Economic Cooperation and Development (OECD), Oct. 2012, pp. 1, 4–39. Available at: `http://search.proquest.com.libproxy.ucl.ac.uk/docview/1321811377/abstract/1085F60A07244EFDPQ/1`.

[217] Goodwin, P. and K. Van Dender. 'Peak Car' — Themes and Issues. *Transport Reviews* 33.3 (May 1, 2013), pp. 243–254.

[218] Van Wee, B. Peak car: The first signs of a shift towards ICT-based activities replacing travel? A discussion paper. *Transport Policy* 42 (Aug. 2015), pp. 1–3.

[219] Bastian, A., M. Börjesson, and J. Eliasson. Explaining "peak car" with economic variables. *Transportation Research Part A: Policy and Practice* 88 (June 2016), pp. 236–250.

[220] Pucher, J., R. Buehler, and M. Seinen. Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A: Policy and Practice* 45.6 (July 2011), pp. 451–475.

[221] Shared-Use Mobility Center (SUMC). *Shared Mobility and the Transformation of Public Transit*. TCRP J-11/TASK 21. American Public Transportation Association, Mar. 2016. Available at: `http://www.trb.org/main/blurbs/174117.aspx`.

[222] Cervero, R. Traffic Impacts of Variable Pricing on the San Francisco-Oakland Bay Bridge, California. *Transportation Research Record: Journal of the Transportation Research Board* 2278.-1 (2012), pp. 145–152.

[223] Barnes, I. C., K. T. Frick, E. Deakin, and A. Skabardonis. Impact of Peak and Off-Peak Tolls on Traffic in San Francisco-Oakland Bay Bridge Corridor in California. *Transportation Research Record: Journal of the Transportation Research Board* 2297.-1 (2012), pp. 73–79.

[224] Pogash, C. Gentrification Spreads an Upheaval in San Francisco's Mission District. *The New York Times* (May 22, 2015).

[225] Pucher, J., J. Dill, and S. Handy. Infrastructure, programs, and policies to increase bicycling: An international review. *Preventive Medicine* 50 (2010), S106–S125.

[226] Hood, J., G. Erhardt, C. Frazier, and A. Schenk. Estimating Emissions Benefits of Bicycle Facilities with Stand-Alone Software Tools. *Transportation Research Record* 2430 (2014), pp. 124–132.

[227] Travis Kalanick. *Uber's Founding*. Dec. 22, 2010. [ONLINE]: `https://newsroom.uber.com/ubers-founding/` (visited on 04/14/2016).

[228] Balcombe, R., R Mackett, N Paulley, J Preston, J Shires, H Titheridge, M Wardman, and P White. *The demand for public transport: a practical guide*. TRL Report 593. TRL Limited, 2004. Available at: `http://www.demandforpublictransport.co.uk/`.

[229] Playfair, W. *The Commercial and Political Atlas*. London: Corry, 1786.

[230] Yau, N. *Bivariate Area Charts in R*. Oct. 1, 2015. [ONLINE]: http://flowingdata.com/2015/10/01/bivariate-area-charts-in-r/ (visited on 04/14/2016).

[231] Spence, I. William Playfair and the psychology of graphs. In: *Proceedings of the American Statistical Association, Section on Statistical Graphics*. Citeseer, 2006, pp. 2426–2436.

[232] Wainer, H. and Spence, I. Graphical Presentation of Longitudinal Data. In: *Encyclopedia of statistics in behavioral science*. Ed. by Everitt, B. and D. C. Howell. Hoboken, N.J: John Wiley & Sons, 2005.

[233] Fox, J. B. Temporal transferability of mode-destination choice models. PhD Thesis. University of Leeds, 2015.

[234] Nicolaisen, M. S. and P. Naess. Roads to nowhere: The accuracy of travel demand forecasts for do-nothing alternatives. *Transport Policy* 37 (2015), pp. 57–63.

[235] Higgins, T. J. and W. L. Johnson. Evaluating transportation programs: neglected principles. *Transportation* 26.4 (1999), pp. 323–336.

[236] Parker, K. M., J. Rice, J. Gustat, J. Ruley, A. Spriggs, and C. Johnson. Effect of Bike Lane Infrastructure Improvements on Ridership in One New Orleans Neighborhood. *Annals of Behavioral Medicine* 45.1 (2013), pp. 101–107.

[237] Arup. *Lane Cove Tunnel : post-opening traffic evaluation : before and after study*. Roads and Traffic Authority, 2009. Available at: http://trid.trb.org/view/1150952.

[238] Erhardt, G. D., S. Patil, T. Light, F. Tsang, P. Burge, P. Sorensen, and M. Zmud. Understanding the Potential of Variable Tolling to Smooth Congestion on Downstream Facilities: Applications of a Joint Time-of-Day and Route Choice Model. *Transportation Research Record* 2563 (2016), pp. 29–39.

[239] London Transport Planning Department. *London Transport Traffic Trends 1971-1990*. Research Report R273. Feb. 1993.

[240] Wardman, M. Review and meta-analysis of U.K. time elasticities of travel demand. *Transportation* 39.3 (Aug. 30, 2011), pp. 465–490.

[241] Graham, D. J., A. Crotte, and R. J. Anderson. A dynamic panel analysis of urban metro demand. *Transportation Research Part E: Logistics and Transportation Review* 45.5 (Sept. 2009), pp. 787–794.

[242] Litman, T. *Understanding transport demands and elasticities*. Victoria Transport Policy Institute, 2013. Available at: `http://www.vtpi.org/elasticities.pdf`.

[243] Wardman, M. Price Elasticities of Surface Travel Demand A Meta-analysis of UK Evidence. *Journal of Transport Economics and Policy (JTEP)* 48.3 (2014), pp. 367–384.

[244] Lock, O. and G. Erhardt. Keeping Track—The Fusion of Large, Automatically Collected Transport Data in Capturing Long-Term System Change. In: Australian Institute of Traffic Planning and Management (AITPM) National Traffic and Transport Conference. Brisbane, Australia, 2015.

[245] Gordon, R. BART considers raising fares, cutting service. *San Francisco Chronicle* (Apr. 10, 2009).

[246] Gordon, R. Board passes BART fare increase, service cuts. *San Francisco Chronicle* (June 12, 2009).

[247] Freeman, D., W. Wei, and G. D. Gosling. *Case Study Report: San Francisco International Airport BART Extension*. Working Paper. Mineta Transportation Institute, San José State University, 2012. Available at: `https://transweb.sjsu.edu/PDFs/research/2503/2503_cases/2503-cs7-SFO-BART-Extension.pdf`.

[248] Murphy, D. BART's Peninsula Line Falls Short of Hopes: Competition from cheaper Baby Bullet trains could be hurting ridership on extension#. *San Francisco Chronicle* (June 30, 2005).

[249] West, R. and P. Herhold. Bay Area Rapid Transit's San Francisco, California, International Airport Station: Assessment of Transit Patronage and Revenue Forecasts. *Transportation Research Record: Journal of the Transportation Research Board* 1895 (2004), pp. 38–45.

[250] Aweiss, A. Bay Area Rapid Transit System San Francisco Extension: Early Findings and Some Tentative Conclusions. In: Transportation Research Board Annual Meeting. Washington, D.C., 2006.

[251] Mason, J. I. Global visions and urban infrastructure analyzing the Bay Area Rapid Transit (BART) extension to San Francisco Airport (SFO). PhD Thesis. University of California, Berkeley, 2008.

[252] San Francisco Municipal Transportation Agency. *SFMTA Strategic Plan: Fiscal Year 2013 - Fiscal Year 2018*. Available at: `https://www.sfmta.com/sites/default/files/pdfs/FY%202013%20-%20FY%202018%20SFMTA%20Strategic%20Plan.pdf`.

[253] Nelson\Nygaard Consulting Associates, Inc. *Municipal Transportation Quality Review: Fiscal Years 2013-2014*. San Francisco Municipal Transportation Agency, 2015. Available at: `https://www.sfmta.com/sites/default/files/reports/2015/SFMTA%20Quality%20Review%202013-14%20FINAL.pdf`.

[254] San Francisco Municipal Transportation Agency. *5L Fulton Limited Pilot Project*. Oct. 20–2013. [ONLINE]: `http://sfmta.com/projects-planning/projects/5l-fulton-limited-pilot-project` (visited on 09/29/2014).

[255] City and County of San Francisco Planning Department. *Transit Effectiveness Project Draft Environmental Impact Report*. Case No. 2011.0558E, State Clearinghouse No. 2011112030. July 10, 2013.

[256] San Francisco Municipal Transportation Agency. *5L Fulton Pilot Final Report*. June 3, 2015. Available at: `https://www.sfmta.com/sites/`

default/files/agendaitems/2015/9-1-15%20Item%2011%205-Fulton%20Report%20-%20accessible.pdf.

[257] Cabanatuan, M. BART can't keep pace with rising 'crush loads'. *San Francisco Chronicle* (Apr. 13, 2015).

[258] Cabanatuan, M. BART riders get perks? There's a catch. *San Francisco Chronicle* (Mar. 3, 2016).

[259] *Urban Engines*. [ONLINE]: http://www.urbanengines.com/ (visited on 04/09/2016).

[260] Merugu, D., B. S. Prabhakar, and N. Rama. An incentive mechanism for decongesting the roads: A pilot program in bangalore. In: *Proceedings of the ACM NetEcon Workshop*. Citeseer, 2009.

[261] Zhu, C., J. S. Yue, C. V. Mandayam, D. Merugu, H. K. Abadi, and B. Prabhakar. Reducing road congestion through incentives: a case study. In: Transportation Research Board Annual Meeting. Washington, D.C., 2015.

[262] *Oakland International Airport: BART to OAK*. [ONLINE]: http://www.oaklandairport.com/bart_airbart.shtml (visited on 04/09/2016).

[263] Federal Transit Administration. *Record of Decision: Bay Area Rapid Transit District Warm Springs Extension Project*. Oct. 24, 2006.

[264] San Francisco Bay Area Rapid Transit District. *Construction Notification #79*. Nov. 2015.

[265] Hunter, T., P. Abbeel, and A. Bayen. The Path Inference Filter: Model-Based Low-Latency Map Matching of Probe Vehicle Data. *IEEE Transactions on Intelligent Transportation Systems* 15.2 (2014), pp. 507–529.

[266] *sfcta/NetworkWrangler*. [ONLINE]: https://github.com/sfcta/NetworkWrangler (visited on 04/19/2016).

[267] *TomTom Maps – Digital Map for Map Based Applications*. [ONLINE]: `http://www.tomtom.com/en_gb/licensing/products/maps/` (visited on 04/19/2016).

[268] Nie, Y., Q. Li, M. Ghamami, and J. Ma. *Urban Travel Reliability Analysis with Consumer GPS Data*. Final Report for SHRP 2 Reliability IDEA Project L15D SHRP 2 Reliability IDEA Program. Transportation Research Board, Oct. 2013.

[269] Bell, M. G. H. The Estimation of an Origin-Destination Matrix form Traffic Counts. *Transportation Science* 17.2 (1983), p. 198.

[270] Hazelton, M. L. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice* 37.10 (2003), pp. 811–822.

[271] Alexander, L., S. Jiang, M. Murga, and M. C. González. Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation Research Part C: Emerging Technologies* 58, Part B (2015), pp. 240–250.

[272] Fox, J., B. Patruni, and A. Daly. *Sydney Strategic Model Re-estimation*. RR 1130. RAND Europe, 2015. Available at: `www.rand.org/t/RR1130`.

[273] Kitamura, R. Panel analysis in transportation planning: An overview. *Transportation Research Part A: General* 24.6 (1990), pp. 401–415.

[274] Chapulut N, J., J. Taroux P, and E. Mange. The new ex-post evaluation methods for large projects in France. In: European Transport Conference. Strassourg, France, 2005.

[275] Bonnafous, A. Permanent Observatories as Tools for Ex-Post Assessment: The French Case Study. In: Roundtable on Ex-post Assessment of Transport Investments and Policy Interventions. Paris: The International Transport Forum at the OECD, Sept. 15–16, 2014.

[276] Vij, A. and K. Shankari. When is big data big enough? Implications of using GPS-based surveys for travel demand analysis. *Transportation Research Part C: Emerging Technologies* 56 (2015), pp. 446–462.

[277] *Cubic Corporation — Global Defense — Transportation Systems*. [ONLINE]: `http://www.cubic.com/` (visited on 04/22/2016).

[278] *AirSage - Home Page*. [ONLINE]: `http://www.airsage.com/` (visited on 04/22/2016).

[279] Ory, D. Travel Modeling as a Science: A Straw Man Proposal. In: *Transportation Research Board Annual Meeting*. Transportation Research Board Annual Meeting. Washington, D.C., 2016.

[280] *Travel modeling as a science - A collaborative, evolving, proposal to improve travel modeling research and practice*. [ONLINE]: `https://github.com/MetropolitanTransportationCommission/travel-modeling-as-science` (visited on 04/22/2016).

Appendix A

Published Works

As declared in the preamble, the work contained in my thesis is my own, and I have indicated where work has been derived from other sources. In that spirit, this appendix lists several related works I have published or presented elsewhere during the course of this research. In cases where those works were published with co-authors, it describes the contribution of each author.

The first paper is based on and closely follows Chapter 3. This work was originally presented at the Workshop on Big Data and Urban Informatics at the University of Illinois at Chicago, Chicago, Illinois in August 2014, and appears in the proceedings from that conference as:

> Erhardt, G.D., Lock, O., Arcaute, E., Batty, M. (2015). "A Big Data Mashing Tool for Measuring Transit System Performance". In *Proceedings of NSF Sponsored Workshop on Big Data and Urban Informatics*, University of Illinois at Chicago, Chicago, Illinois.

After a peer-review process and subsequent changes, the paper was accepted for publication in a book edited by the organisers of that conference. The forthcoming volume is:

> Erhardt, G.D., Lock, O., Arcaute, E., Batty, M. (in-press). "A Big Data Mashing Tool for Measuring Transit System Performance". In *Seeing Cities Through Big Data - Research, Methods and Applications in Urban Informatics*, edited by V. Thakuriah, N. Tilahun, and M. Zellner, Springer, forthcoming.

The contribution of each co-author is: I wrote the document and completed the development of the data processing and reporting tools. Oliver Lock prepared Figures 2, 3 and 4, with Figure 2 based on a visualisation tool developed for his MRes thesis. Elsa Arcaute and Michael Batty provided guidance, direction and review.

The second conference paper is referenced in Chapter 3, and with two of the figures coming from the paper and attributed to the paper.

> Lock, O., Erhardt, G.D. (2015). "Keeping Track—The Fusion of Large, Automatically Collected Transport Data in Capturing Long-Term System Change", paper presented at the 2015 Australian Institute of Traffic Planning and Management (AITPM) National Traffic and Transport Conference, Brisbane, Australia.

This work was written by Oliver Lock and is based on his MRes thesis [117]. Oliver's thesis uses data output from the data fusion tool and to visualise the differences. I provided data for input to Oliver's visualization tools, and supervised his thesis and the subsequent paper.

A third paper is based on Chapter 2. It is:

> Erhardt, G.D. (in-press) "How smart is your smart card? Evaluating transit smart card data with privacy restrictions and limited penetration rates". In *Transportation Research Record*, No. 2544.

This paper was presented as a poster at the 2016 Transportation Research Board Annual Meeting in Washington, D.C. An earlier version appears in the conference pre-prints. It was revised based on reviewer comments, and accepted for publication after a re-review. Chapter 2 is a longer version that incorporates materials both from the pre-print and the publication version.

Three invited presentations were given that draw from the content of this theses. They include:

> Erhardt, G.D. (2015) "Looking Forward and Looking Back: Activity-Based Models and Data Fusion for Improved Travel Forecasts", pre-

sented at the Department of Civil Engineering, University of Kentucky, Lexington, Kentucky.

Erhardt, G.D. (2015) “Back to the Data: From Data-Driven Travel Models to Theory-Driven Travel Models and Back”, presented at the Urban Big Data Centre, University of Glasgow, United Kingdom.

Erhardt, G.D. (2015) “Activity-Based Travel Models and Big Data as Tools for Complementing and Extending UK Research Strengths”, presented at the Department of Civil Engineering, University of Kentucky, Lexington, Kentucky.

Each of these presentations draws from the broad themes described in the introduction, and presents selected materials from Chapter 3 and Chapter 4, in combination with other material.

During the course of this research, I have published or presented several items not related to the content of this thesis. These include:

Erhardt, G.D., Patil, S., Light, T., Tsang, F., Burge, P., Sorenson, P, Zmud, M. (2016). “Understanding the Potential of Variable Tolling to Smooth Congestion on Downstream Facilities: Applications of a Joint Time-of-Day and Route Choice Model”, *Transportation Research Record*, No. 2563.

Light, T., Patil, S., Erhardt, G.D., Tsang, F., Burge, P., Sorensen, P., Zmud, M. (2016). “The Impact of Adopting Time-of-Day Tolling: Case Study of 183A in Austin, Texas”, RAND Corporation, RR-969-CTRMA, Santa Monica, CA.

Erhardt, G.D. (2015). “Using Stated Preference Data and Choice Models to Measure the Impact of Time-of-Day Tolling”, presented at the Transport Modellers’ Forum, London, United Kingdom.

Hood, J., Erhardt, G., Frazier, C., Schenk, A. (2014). “Estimating Emissions Benefits of Bicycle Facilities with Stand-Alone Software Tools”, *Transportation Research Record*, No. 2430.

Erhardt, G.D. (2014). "Microsimulation in Activity-Based Travel Models: Motivation, Stochastic Variation and Opportunities", presented at the Activity-Based Modelling and Appraisal Workshop, University of Hertfordshire, United Kingdom.

Erhardt, G., Hood, J., Frazier, C., Schenk, A. (2013). "Using Smartphone Location Data to Estimate the Air Quality Benefits of Bicycle Infrastructure", presented at the Seminars on Land Use, Transport Models and Big Data, The Martin Centre, University of Cambridge, Cambridge, United Kingdom.

Erhardt, G.D., Sall, E.A., Zorn, L., Tischler, D., Alsup, R., Nassir, N. (2013). "Development and Application of a Dynamic Traffic Assignment Model for San Francisco", presented at the 41st European Transport Conference, Frankfurt, Germany.

The first three are based on work completed through my affiliation with RAND Europe, and the remainder are based on past work completed while I was an employee at Parsons Brinckerhoff.

Appendix B

Enumeration of Full Monthly Time Series

This appendix presents a full enumeration of the monthly performance report data, as referenced in Chapter 4.

The MUNI bus performance report data begins on page 356. Performance report data for other transit modes begins on page 360. The drivers of demand data begin on page 372, and the commute mode shares begin on page 384.

MUNI Performance Report Data

	Trend	Jun-2009	Jul-2009	Aug-2009	Sep-2009	Oct-2009	Nov-2009	Dec-2009	Jan-2010	Feb-2010	Mar-2010	Apr-2010	May-2010
Service Provided													
Vehicle Trips		9,183	9,183	9,183	9,183	9,183	9,183	8,841	8,841	8,841	8,841		8,092
Service Miles		57,751	57,750	57,751	57,751	57,751	57,751	57,627	57,635	57,627	57,627		52,045
Ridership													
Boardings		525,410	510,974	514,078	548,704	528,304	526,566	444,795	478,228	528,024	532,370		499,333
Rear-Door Boardings		1,417	1,642	1,439	1,751	1,612	1,522	1,666	2,115	1,869	1,797		1,419
Passenger Miles		1,024,566	1,004,439	1,010,874	1,073,160	1,030,481	1,026,591	844,146	948,769	1,051,733	1,074,877		1,016,256
Passenger Hours		122,177	119,910	122,039	130,175	124,568	123,855	103,690	112,745	125,749	128,145		122,830
Wheelchairs Served		1,006	1,037	1,022	1,069	1,020	1,130	893	956	1,022	1,146		1,071
Bicycles Served		1,877	1,745	1,685	1,929	1,606	1,522	1,143	1,250	1,467	1,191		1,469
Level-of-Service													
Average Run Speed (mph)		11.01	10.98	10.94	10.97	10.91	10.89	10.84	10.86	10.92	10.82		10.52
Average Total Speed (mph)		8.81	8.80	8.74	8.69	8.73	8.72	8.70	8.68	8.67	8.61		8.37
Average Dwell Time per Stop (min)		0.19	0.19	0.20	0.20	0.20	0.20	0.20	0.21	0.21	0.21		0.22
Average Scheduled Headway (min)		13.84	13.84	13.84	13.84	13.84	13.84	13.19	13.19	13.19	13.19		14.34
Average Full Fare ($)		$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00		$2.00
Average Distance Traveled per Passenger (mi)		1.95	1.97	1.97	1.96	1.95	1.95	1.90	1.98	1.99	2.02		2.04
Average In-Vehicle Time per Passenger (min)		13.95	14.08	14.24	14.23	14.15	14.11	13.99	14.15	14.29	14.44		14.76
Average Wait Time per Passenger (min)		5.30	5.35	5.37	5.31	5.30	5.26	5.19	5.10	5.12	5.11		5.66
Reliability													
Percent of Vehicles Arriving On-Time (-1 to +5 min)		66.5%	66.7%	65.0%	61.2%	61.5%	62.3%	63.8%	60.8%	61.8%	62.3%		62.5%
Average Waiting Delay per Passenger (min)		2.68	2.73	3.15	3.39	3.23	3.03	2.62	2.95	2.72	2.85		2.90
Average Arrival Delay per Passenger (min)		2.18	2.22	2.61	2.85	2.71	2.51	2.09	2.43	2.23	2.35		2.36
Crowding													
Average Volume-Capacity Ratio		0.43	0.42	0.42	0.44	0.43	0.43	0.39	0.40	0.45	0.45		0.46
Percent of Trips with V/C > 0.85		7.6%	6.9%	7.5%	8.8%	8.5%	7.9%	4.7%	5.8%	8.3%	8.8%		8.4%
Observations & Error													
Number of Days		12	22	21	22	22	20	18	20	19	5		15
Days with Observations		12	22	21	22	22	20	18	20	19	5		15
Percent of Trips Observed		19.5%	19.5%	18.3%	17.2%	16.3%	16.6%	17.3%	19.0%	20.0%	21.2%		22.4%
Measurement Error (ON/OFF-1)		-0.2%	-0.3%	0.0%	-0.3%	-0.2%	-0.2%	-1.1%	-1.2%	0.1%	0.3%		0.2%
Weighting Error (SERVMILES/SERVMILES_S-1)		0.6%	0.3%	0.0%	0.4%	0.0%	0.5%	-4.6%	-0.6%	-0.4%	-0.4%		0.4%

MUNI Performance Report Data

	Jun-2010	Jul-2010	Aug-2010	Sep-2010	Oct-2010	Nov-2010	Dec-2010	Jan-2011	Feb-2011	Mar-2011	Apr-2011	May-2011	Jun-2011	Jul-2011	Aug-2011
Service Provided															
Vehicle Trips	8,092	8,092	8,092	8,273	8,273	8,273	8,273	8,368	8,573	8,573	8,573	8,573	8,594	8,606	8,606
Service Miles	52,045	52,045	52,045	54,548	54,548	54,548	54,548	54,677	55,167	55,168	55,167	55,167	55,330	55,422	55,422
Ridership															
Boardings	487,999	493,871	497,995	517,352	518,161	491,770	457,325	493,445	493,509	493,512	516,458	521,283	499,697	492,836	512,709
Rear-Door Boardings	1,920	1,434	1,113	1,931	2,297	1,823	1,771	22,214	66,987	66,501	75,977	78,852	74,128	72,516	79,516
Passenger Miles	997,802	1,008,410	1,018,664	1,067,811	1,058,669	1,007,039	928,499	1,001,662	1,002,441	1,001,781	1,050,546	1,058,980	1,021,833	1,004,043	1,042,973
Passenger Hours	119,133	121,429	122,614	129,669	129,443	123,966	114,440	121,433	120,989	119,989	126,717	128,801	124,540	122,887	128,975
Wheelchairs Served	1,137	1,236	1,305	1,242	1,189	1,229	1,079	1,147	1,199	1,146	1,302	1,297	1,286	1,308	1,235
Bicycles Served	1,595	1,647	2,603	2,387	2,356	2,191	1,322	1,918	1,920	1,576	1,826	1,667	1,640	1,621	1,651
Level-of-Service															
Average Run Speed (mph)	10.63	10.62	10.60	10.59	10.58	10.56	10.53	10.68	10.57	10.63	10.65	10.65	10.66	10.59	10.60
Average Total Speed (mph)	8.46	8.45	8.43	8.43	8.44	8.46	8.50	8.54	8.46	8.52	8.47	8.47	8.50	8.46	8.44
Average Dwell Time per Stop (min)	0.22	0.22	0.22	0.22	0.22	0.21	0.20	0.21	0.21	0.21	0.22	0.22	0.22	0.22	0.22
Average Scheduled Headway (min)	14.34	14.34	14.34	13.86	13.86	13.86	13.86	13.81	13.72	13.72	13.72	13.72	13.71	13.71	13.71
Average Full Fare ($)	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00
Average Distance Traveled per Passenger (mi)	2.04	2.04	2.05	2.06	2.04	2.05	2.03	2.03	2.03	2.03	2.03	2.03	2.04	2.04	2.03
Average In-Vehicle Time per Passenger (min)	14.65	14.75	14.77	15.04	14.99	15.12	15.01	14.77	14.71	14.59	14.72	14.83	14.95	14.96	15.09
Average Wait Time per Passenger (min)	5.65	5.67	5.61	5.39	5.37	5.29	5.32	5.30	5.36	5.36	5.37	5.34	5.38	5.40	5.37
Reliability															
Percent of Vehicles Arriving On-Time (-1 to +5 min)	63.2%	63.0%	63.8%	63.8%	62.3%	61.5%	58.8%	66.8%	66.0%	66.8%	65.2%	64.8%	63.7%	64.3%	63.0%
Average Waiting Delay per Passenger (min)	2.53	2.49	2.64	2.95	3.27	3.34	3.77	2.57	2.76	2.52	3.00	2.84	3.11	3.05	3.22
Average Arrival Delay per Passenger (min)	2.04	1.99	2.13	2.41	2.75	2.80	3.19	2.07	2.25	2.03	2.45	2.30	2.56	2.50	2.65
Crowding															
Average Volume-Capacity Ratio	0.45	0.46	0.46	0.47	0.47	0.44	0.42	0.44	0.44	0.44	0.46	0.47	0.45	0.45	0.47
Percent of Trips with V/C > 0.85	7.7%	8.6%	8.6%	9.3%	9.6%	8.5%	6.1%	7.3%	7.2%	6.7%	8.4%	8.9%	8.3%	7.9%	9.8%
Observations & Error															
Number of Days	22	21	5	18	21	21	23	19	19	23	21	21	22	20	23
Days with Observations	22	21	5	18	21	21	23	19	19	23	21	21	22	20	23
Percent of Trips Observed	22.2%	22.1%	22.7%	22.3%	21.9%	21.1%	19.7%	21.5%	20.6%	22.2%	20.9%	19.6%	21.9%	21.7%	23.2%
Measurement Error (ON/OFF-1)	0.3%	0.4%	0.4%	0.0%	-0.2%	-0.3%	-0.5%	-0.2%	-0.2%	-0.2%	-0.2%	-0.3%	-0.5%	-0.6%	-0.5%
Weighting Error (SERVMILES/SERVMILES_S-1)	0.3%	0.3%	0.7%	0.0%	-0.1%	-0.4%	0.0%	0.8%	0.8%	0.8%	0.7%	0.7%	0.6%	0.1%	0.5%

MUNI Performance Report Data

	Sep-2011	Oct-2011	Nov-2011	Dec-2011	Jan-2012	Feb-2012	Mar-2012	Apr-2012	May-2012	Jun-2012	Jul-2012	Aug-2012	Sep-2012	Oct-2012	Nov-2012
Service Provided															
Vehicle Trips	8,606	8,443	8,311	8,311	8,360	8,452	8,452	8,452	8,452	8,443	8,441	8,441	8,441	8,433	8,433
Service Miles	55,422	55,466	55,501	55,501	55,668	55,976	55,975	55,975	55,975	56,297	56,405	56,405	56,405	56,203	56,203
Ridership															
Boardings	531,730	515,078	491,879	465,621	470,854	499,611	479,422	500,787	489,294	486,691	494,746	515,666	525,409	520,808	488,054
Rear-Door Boardings	88,555	88,982	82,579	77,781	77,553	82,429	75,226	81,096	78,913	83,573	120,387	142,237	160,075	168,279	160,675
Passenger Miles	1,080,306	1,050,869	995,113	940,533	958,063	1,050,758	1,001,621	1,051,472	1,023,378	1,019,872	1,031,961	1,070,092	1,087,250	1,077,324	1,007,825
Passenger Hours	132,626	127,252	120,909	114,541	116,703	125,363	120,752	126,172	122,859	123,773	125,096	131,156	133,096	131,267	121,702
Wheelchairs Served	1,384	1,254	1,273	1,333	1,275	1,270	1,205	1,269	1,310	1,341	1,358	1,392	1,408	1,359	1,304
Bicycles Served	1,799	1,818	1,601	1,475	1,501	1,528	1,554	1,652	1,718	1,746	1,823	1,748	1,679	1,675	1,449
Level-of-Service															
Average Run Speed (mph)	10.63	10.66	10.62	10.58	10.53	10.65	10.69	10.76	10.77	10.60	10.64	10.55	10.45	10.54	10.46
Average Total Speed (mph)	8.40	8.46	8.48	8.48	8.42	8.53	8.55	8.57	8.57	8.44	8.46	8.40	8.33	8.39	8.40
Average Dwell Time per Stop (min)	0.23	0.22	0.22	0.21	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.22	0.23	0.22	0.21
Average Scheduled Headway (min)	13.71	13.77	13.82	13.82	13.80	13.77	13.77	13.77	13.77	13.76	13.75	13.75	13.75	13.72	13.72
Average Full Fare ($)	$2.00	$2.00	$2.00	$1.99	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00
Average Distance Traveled per Passenger (mi)	2.03	2.04	2.02	2.02	2.03	2.10	2.09	2.10	2.09	2.10	2.09	2.08	2.07	2.07	2.06
Average In-Vehicle Time per Passenger (min)	14.97	14.82	14.75	14.76	14.87	15.06	15.11	15.12	15.07	15.26	15.17	15.26	15.20	15.12	14.96
Average Wait Time per Passenger (min)	5.45	5.40	5.41	5.36	5.34	5.49	5.50	5.52	5.47	5.42	5.38	5.41	5.40	5.39	5.37
Reliability															
Percent of Vehicles Arriving On-Time (-1 to +5 min)	62.7%	65.2%	64.3%	64.7%	67.3%	66.1%	64.5%	63.6%	62.5%	62.7%	62.4%	60.5%	61.5%	63.0%	63.8%
Average Waiting Delay per Passenger (min)	3.22	2.90	3.13	3.01	2.77	2.74	2.72	2.75	2.70	2.84	2.83	3.05	3.13	2.81	2.39
Average Arrival Delay per Passenger (min)	2.68	2.38	2.61	2.49	2.22	2.25	2.21	2.25	2.22	2.35	2.32	2.55	2.63	2.30	1.95
Crowding															
Average Volume-Capacity Ratio	0.48	0.46	0.45	0.42	0.43	0.46	0.45	0.47	0.46	0.45	0.45	0.47	0.47	0.46	0.44
Percent of Trips with V/C > 0.85	9.8%	9.1%	8.2%	6.7%	7.1%	8.7%	7.9%	9.1%	8.8%	8.6%	8.6%	10.7%	10.6%	10.2%	7.8%
Observations & Error															
Number of Days	21	20	21	22	20	20	22	21	22	20	21	23	18	23	21
Days with Observations	21	20	21	22	20	20	22	21	22	20	21	23	18	23	21
Percent of Trips Observed	24.6%	24.8%	24.5%	23.5%	23.4%	22.7%	19.9%	19.0%	17.7%	20.9%	20.4%	20.4%	20.2%	21.5%	22.0%
Measurement Error (ON/OFF-1)	-0.3%	0.0%	0.2%	0.3%	0.3%	0.4%	-0.4%	-0.4%	-0.5%	0.2%	0.3%	0.3%	0.3%	0.2%	0.4%
Weighting Error (SERVMILES/SERVMILES_S-1)	0.4%	0.6%	-0.2%	-0.3%	-0.9%	0.4%	0.3%	0.5%	-0.1%	-0.6%	-0.5%	-0.4%	-0.8%	-0.2%	-0.9%

MUNI Performance Report Data

	Dec-2012	Jan-2013	Feb-2013	Mar-2013	Apr-2013	May-2013	Jun-2013	Jul-2013	Aug-2013	Sep-2013	Oct-2013	Nov-2013
Service Provided												
Vehicle Trips	8,433	8,433	8,433	8,433	8,436	8,436	8,436	8,270	8,270	8,270	8,321	8,552
Service Miles	56,203	56,203	56,203	56,203	56,030	56,030	56,030	54,615	54,615	54,615	54,928	56,332
Ridership												
Boardings	502,199	474,027	500,723	494,482	517,413	512,374	479,893	451,808	476,756	512,387	503,591	506,678
Rear-Door Boardings	172,096	158,521	174,965	172,326	186,879	185,704	168,920	155,930	170,407	198,520	199,991	211,558
Passenger Miles	1,010,858	979,707	1,031,493	999,961	1,036,135	1,023,445	962,821	911,358	973,717	1,048,689	1,037,768	1,035,228
Passenger Hours	124,270	117,598	124,475	121,056	124,678	123,010	116,160	112,277	118,435	128,214	124,987	124,228
Wheelchairs Served	1,247	1,226	1,281	1,337	1,365	1,502	1,439	1,326	1,372	1,502	1,632	1,714
Bicycles Served	1,291	1,676	1,517	1,576	1,674	1,693	1,570	1,628	1,779	1,724	1,668	1,544
Level-of-Service												
Average Run Speed (mph)	10.54	10.47	10.42	10.47	10.57	10.67	10.64	10.39	10.61	10.50	10.66	10.75
Average Total Speed (mph)	8.45	8.42	8.34	8.37	8.40	8.49	8.47	8.36	8.48	8.37	8.49	8.54
Average Dwell Time per Stop (min)	0.22	0.21	0.22	0.22	0.23	0.22	0.22	0.21	0.22	0.22	0.22	0.23
Average Scheduled Headway (min)	13.72	13.72	13.72	13.72	13.72	13.72	13.72	14.05	14.05	14.05	14.01	13.84
Average Full Fare ($)	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00
Average Distance Traveled per Passenger (mi)	2.01	2.07	2.06	2.02	2.00	2.00	2.01	2.02	2.04	2.05	2.06	2.04
Average In-Vehicle Time per Passenger (min)	14.85	14.88	14.92	14.69	14.46	14.40	14.52	14.91	14.91	15.01	14.89	14.71
Average Wait Time per Passenger (min)	5.26	5.38	5.39	5.39	5.26	5.41	5.40	5.48	5.57	5.53	5.56	5.38
Reliability												
Percent of Vehicles Arriving On-Time (-1 to +5 min)	59.9%	64.8%	65.1%	65.6%	66.5%	66.1%	65.6%	60.1%	61.4%	63.1%	62.3%	61.1%
Average Waiting Delay per Passenger (min)	3.06	2.22	2.43	2.48	2.57	2.55	2.64	3.29	2.96	3.00	2.86	3.14
Average Arrival Delay per Passenger (min)	2.59	1.78	1.99	2.01	2.06	2.08	2.14	2.80	2.45	2.48	2.36	2.62
Crowding												
Average Volume-Capacity Ratio	0.45	0.43	0.45	0.44	0.45	0.45	0.43	0.43	0.45	0.47	0.45	0.43
Percent of Trips with V/C > 0.85	8.5%	7.0%	7.9%	7.6%	7.9%	7.8%	7.0%	7.6%	8.3%	10.0%	9.0%	7.5%
Observations & Error												
Number of Days	20	21	19	20	22	22	19	22	22	20	22	17
Days with Observations	20	21	19	20	22	22	19	22	22	20	22	17
Percent of Trips Observed	18.1%	23.0%	22.3%	19.6%	19.6%	21.1%	20.4%	16.5%	18.1%	20.3%	22.1%	22.3%
Measurement Error (ON/OFF-1)	-0.6%	0.4%	0.4%	-0.2%	-0.1%	-0.1%	0.0%	-0.1%	0.2%	-0.1%	-0.3%	-0.4%
Weighting Error (SERVMILES/SERVMILES_S-1)	-0.3%	-0.7%	-0.8%	-1.4%	-0.1%	-0.7%	-1.5%	-2.5%	-0.4%	0.1%	0.7%	0.3%

Multi-Modal Performance Data

	Source	Temporal Res	Geog Res	Trend	Jan-2001	Feb-2001	Mar-2001	Apr-2001	May-2001	Jun-2001	Jul-2001	Aug-2001	Sep-2001
Monthly Service Miles													
Muni Bus	Transit Stat Summ	FY	System		1,622,750	1,622,750	1,622,750	1,622,750	1,622,750	1,622,750	1,723,500	1,723,500	1,723,500
Muni Cable Car	Transit Stat Summ	FY	System		41,250	41,250	41,250	41,250	41,250	41,250	36,417	36,417	36,417
Muni Rail	Transit Stat Summ	FY	System		394,833	394,833	394,833	394,833	394,833	394,833	454,917	454,917	454,917
BART	Transit Stat Summ	FY	System		4,897,583	4,897,583	4,897,583	4,897,583	4,897,583	4,897,583	4,869,750	4,869,750	4,869,750
Caltrain	Transit Stat Summ	FY	System		325,917	325,917	325,917	325,917	325,917	325,917	464,250	464,250	464,250
Average Weekday Ridership													
Muni Bus	Transit Stat Summ	FY	System								558,507	558,507	558,507
Muni Cable Car	Transit Stat Summ	FY	System								21,531	21,531	21,531
Muni Rail	Transit Stat Summ	FY	System								157,249	157,249	157,249
BART	Transit Stat Summ	FY	System								329,527	329,527	329,527
Caltrain	Transit Stat Summ	FY	System								27,663	27,663	27,663
Average Weekday Ridership													
Muni Bus	APCs/Faregate	Monthly	Stop										
BART	APCs/Faregate	Monthly	Stop		321,181	334,885	340,022	333,794	333,847	329,448	324,566	321,758	321,321
Cash Fare (2010$)													
Muni Bus+Rail	Published Values	Actual	System		$1.25	$1.24	$1.24	$1.23	$1.23	$1.23	$1.23	$1.23	$1.22
Muni Cable Car	Published Values	Actual	System		$2.49	$2.48	$2.48	$2.47	$2.45	$2.45	$2.46	$2.46	$2.45
BART	Published Values	Actual	System		$3.26	$3.25	$3.24	$3.23	$3.22	$3.21	$3.22	$3.22	$3.20
Average Fare (2010$)													
Muni Bus+Rail	Transit Stat Summ	FY/Actual	System		$0.53	$0.53	$0.53	$0.53	$0.52	$0.52	$0.47	$0.47	$0.47
Muni Cable Car	Transit Stat Summ	FY/Actual	System		$0.53	$0.53	$0.53	$0.53	$0.52	$0.52	$1.76	$1.76	$1.75
BART	Transit Stat Summ	FY/Actual	System		$2.56	$2.55	$2.54	$2.53	$2.52	$2.51	$2.44	$2.44	$2.43
Caltrain	Transit Stat Summ	FY/Actual	System		$2.86	$2.85	$2.84	$2.83	$2.82	$2.81	$3.18	$3.18	$3.17
Weekday Service Miles													
Muni Bus	GTFS	Actual	Route/Stop										
Muni Cable Car	GTFS	Actual	Route/Stop										
Muni Rail	GTFS	Actual	Route/Stop										
BART	GTFS	Actual	Route/Stop										
Weekday Service Miles-Extrapolated													
Muni Bus	Stat Summ/GTFS	Monthly	System		59,145	59,145	59,145	59,145	59,145	59,145	62,817	62,817	62,817
Muni Cable Car	Stat Summ/GTFS	Monthly	System		1,946	1,946	1,946	1,946	1,946	1,946	1,718	1,718	1,718
Muni Rail	Stat Summ/GTFS	Monthly	System		14,844	14,844	14,844	14,844	14,844	14,844	17,102	17,102	17,102
BART	Stat Summ/GTFS	Monthly	System		25,527	25,527	25,527	25,527	25,527	25,527	25,382	25,382	25,382

Multi-Modal Performance Data

	Oct-2001	Nov-2001	Dec-2001	Jan-2002	Feb-2002	Mar-2002	Apr-2002	May-2002	Jun-2002	Jul-2002	Aug-2002	Sep-2002	Oct-2002	Nov-2002	Dec-2002	Jan-2003
Monthly Service Miles																
Muni Bus	1,723,500	1,723,500	1,723,500	1,723,500	1,723,500	1,723,500	1,723,500	1,723,500	1,723,500	1,728,917	1,728,917	1,728,917	1,728,917	1,728,917	1,728,917	1,728,917
Muni Cable Car	36,417	36,417	36,417	36,417	36,417	36,417	36,417	36,417	36,417	33,750	33,750	33,750	33,750	33,750	33,750	33,750
Muni Rail	454,917	454,917	454,917	454,917	454,917	454,917	454,917	454,917	454,917	460,917	460,917	460,917	460,917	460,917	460,917	460,917
BART	4,869,750	4,869,750	4,869,750	4,869,750	4,869,750	4,869,750	4,869,750	4,869,750	4,869,750	4,906,750	4,906,750	4,906,750	4,906,750	4,906,750	4,906,750	4,906,750
Caltrain	464,250	464,250	464,250	464,250	464,250	464,250	464,250	464,250	464,250	402,750	402,750	402,750	402,750	402,750	402,750	402,750
Average Weekday Ridership																
Muni Bus	558,507	558,507	558,507	558,507	558,507	558,507	558,507	558,507	558,507	511,466	511,466	511,466	511,466	511,466	511,466	511,466
Muni Cable Car	21,531	21,531	21,531	21,531	21,531	21,531	21,531	21,531	21,531	20,564	20,564	20,564	20,564	20,564	20,564	20,564
Muni Rail	157,249	157,249	157,249	157,249	157,249	157,249	157,249	157,249	157,249	136,629	136,629	136,629	136,629	136,629	136,629	136,629
BART	329,527	329,527	329,527	329,527	329,527	329,527	329,527	329,527	329,527	313,577	313,577	313,577	313,577	313,577	313,577	313,577
Caltrain	27,663	27,663	27,663	27,663	27,663	27,663	27,663	27,663	27,663	25,023	25,023	25,023	25,023	25,023	25,023	25,023
Average Weekday Ridership																
Muni Bus																
BART	315,571	317,825	292,391	300,938	307,748	305,537	311,873	304,468	302,730	300,921	296,473	309,640	303,796	294,360	282,622	288,638
Cash Fare (2010$)																
Muni Bus+Rail	$1.23	$1.23	$1.23	$1.23	$1.23	$1.22	$1.21	$1.21	$1.21	$1.21	$1.21	$1.20	$1.20	$1.20	$1.21	$1.50
Muni Cable Car	$2.45	$2.46	$2.47	$2.46	$2.45	$2.44	$2.43	$2.43	$2.42	$2.42	$2.41	$2.41	$2.41	$2.41	$2.41	$2.40
BART	$3.22	$3.22	$3.23	$3.23	$3.21	$3.20	$3.18	$3.18	$3.18	$3.17	$3.16	$3.16	$3.15	$3.15	$3.16	$3.31
Average Fare (2010$)																
Muni Bus+Rail	$0.47	$0.47	$0.47	$0.47	$0.47	$0.47	$0.46	$0.46	$0.46	$0.45	$0.45	$0.45	$0.45	$0.45	$0.45	$0.56
Muni Cable Car	$1.76	$1.76	$1.77	$1.76	$1.76	$1.75	$1.74	$1.74	$1.74	$1.80	$1.79	$1.79	$1.79	$1.79	$1.79	$1.78
BART	$2.44	$2.45	$2.45	$2.45	$2.44	$2.43	$2.41	$2.41	$2.41	$2.41	$2.40	$2.39	$2.39	$2.39	$2.40	$2.51
Caltrain	$3.18	$3.18	$3.20	$3.19	$3.18	$3.16	$3.14	$3.14	$3.14	$3.80	$3.79	$3.78	$3.78	$3.78	$3.79	$3.77
Weekday Service Miles																
Muni Bus																
Muni Cable Car																
Muni Rail																
BART																
Weekday Service Miles-Extrapolated																
Muni Bus	62,817	62,817	62,817	62,817	62,817	62,817	62,817	62,817	62,817	63,015	63,015	63,015	63,015	63,015	63,015	63,015
Muni Cable Car	1,718	1,718	1,718	1,718	1,718	1,718	1,718	1,718	1,718	1,592	1,592	1,592	1,592	1,592	1,592	1,592
Muni Rail	17,102	17,102	17,102	17,102	17,102	17,102	17,102	17,102	17,102	17,328	17,328	17,328	17,328	17,328	17,328	17,328
BART	25,382	25,382	25,382	25,382	25,382	25,382	25,382	25,382	25,382	25,575	25,575	25,575	25,575	25,575	25,575	25,575

Multi-Modal Performance Data

	Feb-2003	Mar-2003	Apr-2003	May-2003	Jun-2003	Jul-2003	Aug-2003	Sep-2003	Oct-2003	Nov-2003	Dec-2003	Jan-2004	Feb-2004	Mar-2004	Apr-2004	May-2004
Monthly Service Miles																
Muni Bus	1,728,917	1,728,917	1,728,917	1,728,917	1,728,917	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583	1,687,583
Muni Cable Car	33,750	33,750	33,750	33,750	33,750	37,750	37,750	37,750	37,750	37,750	37,750	37,750	37,750	37,750	37,750	37,750
Muni Rail	460,917	460,917	460,917	460,917	460,917	471,333	471,333	471,333	471,333	471,333	471,333	471,333	471,333	471,333	471,333	471,333
BART	4,906,750	4,906,750	4,906,750	4,906,750	4,906,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750	5,197,750
Caltrain	402,750	402,750	402,750	402,750	402,750	430,833	430,833	430,833	430,833	430,833	430,833	430,833	430,833	430,833	430,833	430,833
Average Weekday Ridership																
Muni Bus	511,466	511,466	511,466	511,466	511,466	550,992	550,992	550,992	550,992	550,992	550,992	550,992	550,992	550,992	550,992	550,992
Muni Cable Car	20,564	20,564	20,564	20,564	20,564	21,637	21,637	21,637	21,637	21,637	21,637	21,637	21,637	21,637	21,637	21,637
Muni Rail	136,629	136,629	136,629	136,629	136,629	143,784	143,784	143,784	143,784	143,784	143,784	143,784	143,784	143,784	143,784	143,784
BART	313,577	313,577	313,577	313,577	313,577	324,993	324,993	324,993	324,993	324,993	324,993	324,993	324,993	324,993	324,993	324,993
Caltrain	25,023	25,023	25,023	25,023	25,023	24,545	24,545	24,545	24,545	24,545	24,545	24,545	24,545	24,545	24,545	24,545
Average Weekday Ridership																
Muni Bus																
BART	295,525	296,030	292,949	289,681	290,310	298,911	302,388	313,682	315,791	308,966	296,470	296,820	305,248	308,214	310,872	310,777
Cash Fare (2010$)																
Muni Bus+Rail	$1.49	$1.48	$1.48	$1.49	$1.48	$1.48	$1.48	$1.47	$1.47	$1.48	$1.48	$1.47	$1.46	$1.45	$1.45	$1.44
Muni Cable Car	$2.38	$2.37	$2.37	$2.38	$2.37	$2.37	$2.36	$3.53	$3.54	$3.55	$3.55	$3.53	$3.51	$3.49	$3.48	$3.46
BART	$3.29	$3.27	$3.27	$3.28	$3.28	$3.27	$3.26	$3.25	$3.25	$3.26	$3.27	$3.58	$3.56	$3.54	$3.53	$3.51
Average Fare (2010$)																
Muni Bus+Rail	$0.55	$0.55	$0.55	$0.55	$0.55	$0.60	$0.60	$0.59	$0.59	$0.60	$0.60	$0.59	$0.59	$0.59	$0.58	$0.58
Muni Cable Car	$1.77	$1.76	$1.76	$1.77	$1.76	$1.65	$1.64	$2.45	$2.45	$2.46	$2.46	$2.45	$2.44	$2.42	$2.41	$2.40
BART	$2.49	$2.48	$2.48	$2.49	$2.48	$2.54	$2.53	$2.53	$2.53	$2.54	$2.54	$2.78	$2.77	$2.75	$2.74	$2.73
Caltrain	$3.74	$3.72	$3.73	$3.73	$3.73	$3.05	$3.04	$3.03	$3.03	$3.04	$3.04	$3.03	$3.01	$2.99	$2.98	$2.97
Weekday Service Miles																
Muni Bus																
Muni Cable Car																
Muni Rail																
BART																
Weekday Service Miles-Extrapolated																
Muni Bus	63,015	63,015	63,015	63,015	63,015	61,508	61,508	61,508	61,508	61,508	61,508	61,508	61,508	61,508	61,508	61,508
Muni Cable Car	1,592	1,592	1,592	1,592	1,592	1,781	1,781	1,781	1,781	1,781	1,781	1,781	1,781	1,781	1,781	1,781
Muni Rail	17,328	17,328	17,328	17,328	17,328	17,720	17,720	17,720	17,720	17,720	17,720	17,720	17,720	17,720	17,720	17,720
BART	25,575	25,575	25,575	25,575	25,575	27,092	27,092	27,092	27,092	27,092	27,092	27,092	27,092	27,092	27,092	27,092

Multi-Modal Performance Data

	Jun-2004	Jul-2004	Aug-2004	Sep-2004	Oct-2004	Nov-2004	Dec-2004	Jan-2005	Feb-2005	Mar-2005	Apr-2005	May-2005	Jun-2005	Jul-2005	Aug-2005	Sep-2005
Monthly Service Miles																
Muni Bus	1,687,583	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,709,500	1,585,083	1,585,083	1,585,083
Muni Cable Car	37,750	34,500	34,500	34,500	34,500	34,500	34,500	34,500	34,500	34,500	34,500	34,500	34,500	36,333	36,333	36,333
Muni Rail	471,333	460,417	460,417	460,417	460,417	460,417	460,417	460,417	460,417	460,417	460,417	460,417	460,417	446,500	446,500	446,500
BART	5,197,750	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,000,333	5,174,083	5,174,083	5,174,083
Caltrain	430,833	459,000	459,000	459,000	459,000	459,000	459,000	459,000	459,000	459,000	459,000	459,000	459,000	513,167	513,167	513,167
Average Weekday Ridership																
Muni Bus	550,992	518,513	518,513	518,513	518,513	518,513	518,513	518,513	518,513	518,513	518,513	518,513	518,513	510,300	510,300	510,300
Muni Cable Car	21,637	19,166	19,166	19,166	19,166	19,166	19,166	19,166	19,166	19,166	19,166	19,166	19,166	21,629	21,629	21,629
Muni Rail	143,784	146,862	146,862	146,862	146,862	146,862	146,862	146,862	146,862	146,862	146,862	146,862	146,862	132,637	132,637	132,637
BART	324,993	329,199	329,199	329,199	329,199	329,199	329,199	329,199	329,199	329,199	329,199	329,199	329,199	343,026	343,026	343,026
Caltrain	24,545	27,487	27,487	27,487	27,487	27,487	27,487	27,487	27,487	27,487	27,487	27,487	27,487	32,291	32,291	32,291
Average Weekday Ridership																
Muni Bus																
BART	308,793	308,190	304,724	323,236	315,753	313,975	294,788	297,524	310,053	310,722	319,649	315,888	314,079	309,940	317,614	336,338
Cash Fare (2010$)																
Muni Bus+Rail	$1.44	$1.44	$1.44	$1.44	$1.43	$1.43	$1.43	$1.43	$1.42	$1.41	$1.40	$1.40	$1.40	$1.39	$1.39	$1.65
Muni Cable Car	$3.45	$3.45	$3.45	$3.44	$3.43	$3.42	$3.44	$3.43	$3.41	$3.38	$3.36	$3.37	$3.36	$3.35	$3.33	$5.48
BART	$3.49	$3.50	$3.50	$3.49	$3.47	$3.47	$3.48	$3.48	$3.46	$3.43	$3.41	$3.41	$3.41	$3.39	$3.38	$3.33
Average Fare (2010$)																
Muni Bus+Rail	$0.58	$0.57	$0.57	$0.57	$0.57	$0.57	$0.57	$0.57	$0.56	$0.56	$0.55	$0.56	$0.56	$0.55	$0.55	$0.65
Muni Cable Car	$2.39	$2.68	$2.68	$2.68	$2.66	$2.66	$2.67	$2.66	$2.65	$2.63	$2.61	$2.61	$2.61	$1.41	$1.40	$2.31
BART	$2.72	$2.70	$2.70	$2.70	$2.68	$2.68	$2.69	$2.68	$2.67	$2.65	$2.63	$2.63	$2.63	$2.70	$2.69	$2.65
Caltrain	$2.96	$2.35	$2.34	$2.34	$2.33	$2.33	$2.33	$2.33	$2.32	$2.30	$2.28	$2.28	$2.28	$3.34	$3.32	$3.28
Weekday Service Miles																
Muni Bus																
Muni Cable Car																
Muni Rail																
BART																
Weekday Service Miles-Extrapolated																
Muni Bus	61,508	62,307	62,307	62,307	62,307	62,307	62,307	62,307	62,307	62,307	62,307	62,307	62,307	57,772	57,772	57,772
Muni Cable Car	1,781	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,627	1,714	1,714	1,714
Muni Rail	17,720	17,309	17,309	17,309	17,309	17,309	17,309	17,309	17,309	17,309	17,309	17,309	17,309	16,786	16,786	16,786
BART	27,092	26,063	26,063	26,063	26,063	26,063	26,063	26,063	26,063	26,063	26,063	26,063	26,063	26,968	26,968	26,968

Multi-Modal Performance Data

	Oct-2005	Nov-2005	Dec-2005	Jan-2006	Feb-2006	Mar-2006	Apr-2006	May-2006	Jun-2006	Jul-2006	Aug-2006	Sep-2006	Oct-2006	Nov-2006	Dec-2006	Jan-2007
Monthly Service Miles																
Muni Bus	1,585,083	1,585,083	1,585,083	1,585,083	1,585,083	1,585,083	1,585,083	1,585,083	1,585,083	1,545,000	1,545,000	1,545,000	1,545,000	1,545,000	1,545,000	1,545,000
Muni Cable Car	36,333	36,333	36,333	36,333	36,333	36,333	36,333	36,333	36,333	39,083	39,083	39,083	39,083	39,083	39,083	39,083
Muni Rail	446,500	446,500	446,500	446,500	446,500	446,500	446,500	446,500	446,500	422,833	422,833	422,833	422,833	422,833	422,833	422,833
BART	5,174,083	5,174,083	5,174,083	5,174,083	5,174,083	5,174,083	5,174,083	5,174,083	5,174,083	5,507,583	5,507,583	5,507,583	5,507,583	5,507,583	5,507,583	5,507,583
Caltrain	513,167	513,167	513,167	513,167	513,167	513,167	513,167	513,167	513,167	520,333	520,333	520,333	520,333	520,333	520,333	520,333
Average Weekday Ridership																
Muni Bus	510,300	510,300	510,300	510,300	510,300	510,300	510,300	510,300	510,300	499,941	499,941	499,941	499,941	499,941	499,941	499,941
Muni Cable Car	21,629	21,629	21,629	21,629	21,629	21,629	21,629	21,629	21,629	19,041	19,041	19,041	19,041	19,041	19,041	19,041
Muni Rail	132,637	132,637	132,637	132,637	132,637	132,637	132,637	132,637	132,637	130,914	130,914	130,914	130,914	130,914	130,914	130,914
BART	343,026	343,026	343,026	343,026	343,026	343,026	343,026	343,026	343,026	362,483	362,483	362,483	362,483	362,483	362,483	362,483
Caltrain	32,291	32,291	32,291	32,291	32,291	32,291	32,291	32,291	32,291	34,867	34,867	34,867	34,867	34,867	34,867	34,867
Average Weekday Ridership																
Muni Bus																
BART	332,172	331,713	304,147	314,749	324,285	317,615	323,465	331,373	332,733	337,259	334,426	345,594	344,855	339,067	323,739	323,871
Cash Fare (2010$)																
Muni Bus+Rail	$1.64	$1.66	$1.66	$1.65	$1.65	$1.64	$1.62	$1.62	$1.61	$1.61	$1.60	$1.61	$1.62	$1.62	$1.62	$1.62
Muni Cable Car	$5.47	$5.52	$5.54	$5.50	$5.49	$5.46	$5.41	$5.38	$5.37	$5.36	$5.35	$5.37	$5.40	$5.41	$5.40	$5.39
BART	$3.33	$3.35	$3.37	$3.46	$3.46	$3.44	$3.41	$3.39	$3.39	$3.38	$3.37	$3.39	$3.40	$3.41	$3.40	$3.39
Average Fare (2010$)																
Muni Bus+Rail	$0.65	$0.65	$0.66	$0.65	$0.65	$0.65	$0.64	$0.64	$0.64	$0.68	$0.68	$0.68	$0.69	$0.69	$0.69	$0.68
Muni Cable Car	$2.31	$2.33	$2.34	$2.32	$2.31	$2.30	$2.28	$2.27	$2.27	$2.25	$2.24	$2.25	$2.26	$2.27	$2.26	$2.26
BART	$2.65	$2.67	$2.68	$2.76	$2.75	$2.74	$2.71	$2.70	$2.69	$2.76	$2.75	$2.77	$2.78	$2.78	$2.78	$2.77
Caltrain	$3.28	$3.30	$3.32	$3.29	$3.28	$3.27	$3.24	$3.22	$3.22	$3.40	$3.39	$3.41	$3.43	$3.43	$3.43	$3.42
Weekday Service Miles																
Muni Bus																
Muni Cable Car																
Muni Rail																
BART																28,707
Weekday Service Miles-Extrapolated																
Muni Bus	57,772	57,772	57,772	57,772	57,772	57,772	57,772	57,772	57,772	56,311	56,311	56,311	56,311	56,311	56,311	56,311
Muni Cable Car	1,714	1,714	1,714	1,714	1,714	1,714	1,714	1,714	1,714	1,843	1,843	1,843	1,843	1,843	1,843	1,843
Muni Rail	16,786	16,786	16,786	16,786	16,786	16,786	16,786	16,786	16,786	15,896	15,896	15,896	15,896	15,896	15,896	15,896
BART	26,968	26,968	26,968	26,968	26,968	26,968	26,968	26,968	26,968	28,707	28,707	28,707	28,707	28,707	28,707	28,707

Multi-Modal Performance Data

	Feb-2007	Mar-2007	Apr-2007	May-2007	Jun-2007	Jul-2007	Aug-2007	Sep-2007	Oct-2007	Nov-2007	Dec-2007	Jan-2008	Feb-2008	Mar-2008	Apr-2008	May-2008
Monthly Service Miles																
Muni Bus	1,545,000	1,545,000	1,545,000	1,545,000	1,545,000	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250	1,567,250
Muni Cable Car	39,083	39,083	39,083	39,083	39,083	39,833	39,833	39,833	39,833	39,833	39,833	39,833	39,833	39,833	39,833	39,833
Muni Rail	422,833	422,833	422,833	422,833	422,833	340,750	340,750	340,750	340,750	340,750	340,750	340,750	340,750	340,750	340,750	340,750
BART	5,507,583	5,507,583	5,507,583	5,507,583	5,507,583	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333	5,582,333
Caltrain	520,333	520,333	520,333	520,333	520,333	559,167	559,167	559,167	559,167	559,167	559,167	559,167	559,167	559,167	559,167	559,167
Average Weekday Ridership																
Muni Bus	499,941	499,941	499,941	499,941	499,941	507,212	507,212	507,212	507,212	507,212	507,212	507,212	507,212	507,212	507,212	507,212
Muni Cable Car	19,041	19,041	19,041	19,041	19,041	20,530	20,530	20,530	20,530	20,530	20,530	20,530	20,530	20,530	20,530	20,530
Muni Rail	130,914	130,914	130,914	130,914	130,914	159,405	159,405	159,405	159,405	159,405	159,405	159,405	159,405	159,405	159,405	159,405
BART	362,483	362,483	362,483	362,483	362,483	384,231	384,231	384,231	384,231	384,231	384,231	384,231	384,231	384,231	384,231	384,231
Caltrain	34,867	34,867	34,867	34,867	34,867	37,000	37,000	37,000	37,000	37,000	37,000	37,000	37,000	37,000	37,000	37,000
Average Weekday Ridership																
Muni Bus																
BART	333,519	342,005	342,994	352,993	349,709	349,717	354,772	364,631	360,197	365,425	347,133	336,321	356,237	355,646	367,603	362,761
Cash Fare (2010$)																
Muni Bus+Rail	$1.61	$1.59	$1.58	$1.57	$1.57	$1.57	$1.57	$1.57	$1.57	$1.56	$1.56	$1.55	$1.55	$1.53	$1.52	$1.51
Muni Cable Car	$5.36	$5.31	$5.28	$5.24	$5.23	$5.23	$5.24	$5.23	$5.22	$5.19	$5.19	$5.17	$5.15	$5.11	$5.08	$5.03
BART	$3.38	$3.34	$3.32	$3.30	$3.30	$3.30	$3.30	$3.29	$3.29	$3.27	$3.27	$3.44	$3.43	$3.40	$3.38	$3.35
Average Fare (2010$)																
Muni Bus+Rail	$0.68	$0.67	$0.67	$0.67	$0.66	$0.62	$0.62	$0.62	$0.62	$0.61	$0.61	$0.61	$0.61	$0.60	$0.60	$0.59
Muni Cable Car	$2.25	$2.23	$2.21	$2.20	$2.19	$3.42	$3.42	$3.42	$3.41	$3.39	$3.39	$3.37	$3.36	$3.33	$3.31	$3.29
BART	$2.76	$2.73	$2.72	$2.70	$2.69	$2.73	$2.74	$2.73	$2.72	$2.71	$2.71	$2.85	$2.84	$2.82	$2.80	$2.78
Caltrain	$3.40	$3.37	$3.35	$3.33	$3.32	$3.36	$3.37	$3.36	$3.35	$3.33	$3.33	$3.32	$3.31	$3.28	$3.26	$3.23
Weekday Service Miles																
Muni Bus																
Muni Cable Car																
Muni Rail																
BART	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707
Weekday Service Miles-Extrapolated																
Muni Bus	56,311	56,311	56,311	56,311	56,311	57,122	57,122	57,122	57,122	57,122	57,122	57,122	57,122	57,122	57,122	57,122
Muni Cable Car	1,843	1,843	1,843	1,843	1,843	1,879	1,879	1,879	1,879	1,879	1,879	1,879	1,879	1,879	1,879	1,879
Muni Rail	15,896	15,896	15,896	15,896	15,896	12,810	12,810	12,810	12,810	12,810	12,810	12,810	12,810	12,810	12,810	12,810
BART	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707

Multi-Modal Performance Data

	Jun-2008	Jul-2008	Aug-2008	Sep-2008	Oct-2008	Nov-2008	Dec-2008	Jan-2009	Feb-2009	Mar-2009	Apr-2009	May-2009	Jun-2009	Jul-2009	Aug-2009	Sep-2009
Monthly Service Miles																
Muni Bus	1,567,250	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,588,917	1,557,917	1,557,917	1,557,917
Muni Cable Car	39,833	28,667	28,667	28,667	28,667	28,667	28,667	28,667	28,667	28,667	28,667	28,667	28,667	28,500	28,500	28,500
Muni Rail	340,750	342,917	342,917	342,917	342,917	342,917	342,917	342,917	342,917	342,917	342,917	342,917	342,917	344,917	344,917	344,917
BART	5,582,333	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,653,583	5,269,833	5,269,833	5,269,833
Caltrain	559,167	574,667	574,667	574,667	574,667	574,667	574,667	574,667	574,667	574,667	574,667	574,667	574,667	547,500	547,500	547,500
Average Weekday Ridership																
Muni Bus	507,212	522,324	522,324	522,324	522,324	522,324	522,324	522,324	522,324	522,324	522,324	522,324	522,324	495,310	495,310	495,310
Muni Cable Car	20,530	21,542	21,542	21,542	21,542	21,542	21,542	21,542	21,542	21,542	21,542	21,542	21,542	22,353	22,353	22,353
Muni Rail	159,405	163,593	163,593	163,593	163,593	163,593	163,593	163,593	163,593	163,593	163,593	163,593	163,593	158,430	158,430	158,430
BART	384,231	379,007	379,007	379,007	379,007	379,007	379,007	379,007	379,007	379,007	379,007	379,007	379,007	357,461	357,461	357,461
Caltrain	37,000	40,066	40,066	40,066	40,066	40,066	40,066	40,066	40,066	40,066	40,066	40,066	40,066	37,745	37,745	37,745
Average Weekday Ridership																
Muni Bus													525,410	510,974	514,078	548,704
BART	372,343	374,967	369,884	379,996	370,502	364,329	351,130	344,786	348,046	346,658	347,237	342,394	339,133	332,368	329,688	353,681
Cash Fare (2010$)																
Muni Bus+Rail	$1.49	$1.49	$1.49	$1.50	$1.51	$1.54	$1.56	$1.55	$1.54	$1.54	$1.53	$1.53	$1.52	$2.03	$2.02	$2.02
Muni Cable Car	$4.98	$4.96	$4.98	$4.98	$5.03	$5.13	$5.19	$5.16	$5.14	$5.13	$5.11	$5.10	$5.05	$5.06	$5.05	$5.05
BART	$3.32	$3.30	$3.31	$3.32	$3.35	$3.42	$3.45	$3.44	$3.42	$3.41	$3.41	$3.40	$3.37	$3.58	$3.58	$3.57
Average Fare (2010$)																
Muni Bus+Rail	$0.59	$0.57	$0.57	$0.57	$0.58	$0.59	$0.60	$0.60	$0.59	$0.59	$0.59	$0.59	$0.58	$0.78	$0.78	$0.78
Muni Cable Car	$3.25	$3.09	$3.10	$3.11	$3.14	$3.20	$3.23	$3.22	$3.20	$3.20	$3.19	$3.18	$3.15	$3.24	$3.23	$3.23
BART	$2.75	$2.75	$2.76	$2.77	$2.79	$2.85	$2.88	$2.87	$2.85	$2.84	$2.84	$2.83	$2.81	$3.10	$3.09	$3.09
Caltrain	$3.20	$3.38	$3.39	$3.40	$3.43	$3.50	$3.54	$3.52	$3.50	$3.50	$3.49	$3.48	$3.45	$4.08	$4.07	$4.07
Weekday Service Miles																
Muni Bus									57,912	57,912	57,912	57,217	57,824	57,223	57,750	57,750
Muni Cable Car									1,352	1,352	1,352	1,348	1,352	1,349	1,352	1,352
Muni Rail									12,892	12,892	12,892	12,655	14,975	16,527	16,710	16,710
BART	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	27,759
Weekday Service Miles-Extrapolated																
Muni Bus	57,122	57,912	57,912	57,912	57,912	57,912	57,912	57,912	57,912	57,912	57,912	57,217	57,824	57,223	57,750	57,750
Muni Cable Car	1,879	1,352	1,352	1,352	1,352	1,352	1,352	1,352	1,352	1,352	1,352	1,348	1,352	1,349	1,352	1,352
Muni Rail	12,810	12,892	12,892	12,892	12,892	12,892	12,892	12,892	12,892	12,892	12,892	12,655	14,975	16,527	16,710	16,710
BART	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	28,707	27,759

Multi-Modal Performance Data

	Oct-2009	Nov-2009	Dec-2009	Jan-2010	Feb-2010	Mar-2010	Apr-2010	May-2010	Jun-2010	Jul-2010	Aug-2010	Sep-2010	Oct-2010	Nov-2010	Dec-2010	Jan-2011
Monthly Service Miles																
Muni Bus	1,557,917	1,557,917	1,557,917	1,557,917	1,557,917	1,557,917	1,557,917	1,557,917	1,557,917	1,477,417	1,477,417	1,477,417	1,477,417	1,477,417	1,477,417	1,477,417
Muni Cable Car	28,500	28,500	28,500	28,500	28,500	28,500	28,500	28,500	28,500	23,917	23,917	23,917	23,917	23,917	23,917	23,917
Muni Rail	344,917	344,917	344,917	344,917	344,917	344,917	344,917	344,917	344,917	486,500	486,500	486,500	486,500	486,500	486,500	486,500
BART	5,269,833	5,269,833	5,269,833	5,269,833	5,269,833	5,269,833	5,269,833	5,269,833	5,269,833	5,278,917	5,278,917	5,278,917	5,278,917	5,278,917	5,278,917	5,278,917
Caltrain	547,500	547,500	547,500	547,500	547,500	547,500	547,500	547,500	547,500	541,833	541,833	541,833	541,833	541,833	541,833	541,833
Average Weekday Ridership																
Muni Bus	495,310	495,310	495,310	495,310	495,310	495,310	495,310	495,310	495,310	491,906	491,906	491,906	491,906	491,906	491,906	491,906
Muni Cable Car	22,353	22,353	22,353	22,353	22,353	22,353	22,353	22,353	22,353	19,893	19,893	19,893	19,893	19,893	19,893	19,893
Muni Rail	158,430	158,430	158,430	158,430	158,430	158,430	158,430	158,430	158,430	161,398	161,398	161,398	161,398	161,398	161,398	161,398
BART	357,461	357,461	357,461	357,461	357,461	357,461	357,461	357,461	357,461	367,505	367,505	367,505	367,505	367,505	367,505	367,505
Caltrain	37,745	37,745	37,745	37,745	37,745	37,745	37,745	37,745	37,745	39,090	39,090	39,090	39,090	39,090	39,090	39,090
Average Weekday Ridership																
Muni Bus	528,304	526,566	444,795	478,228	528,024	532,370		499,333	487,999	493,871	497,995	517,352	518,161	491,770	457,325	493,445
BART	358,409	345,355	322,558	318,826	333,792	329,811	333,866	331,934	329,541	332,894	338,846	354,579	356,545	353,502	323,669	334,045
Cash Fare (2010$)																
Muni Bus+Rail	$2.02	$2.02	$2.02	$2.01	$2.01	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$2.00	$1.99	$1.99	$1.99	$1.98
Muni Cable Car	$5.04	$5.04	$5.05	$5.03	$5.03	$5.01	$5.00	$5.00	$5.00	$5.00	$4.99	$4.99	$4.99	$4.98	$4.97	$4.95
BART	$3.57	$3.57	$3.57	$3.56	$3.56	$3.55	$3.54	$3.54	$3.54	$3.54	$3.54	$3.53	$3.53	$3.53	$3.52	$3.51
Average Fare (2010$)																
Muni Bus+Rail	$0.78	$0.78	$0.78	$0.78	$0.78	$0.77	$0.77	$0.77	$0.77	$0.80	$0.80	$0.80	$0.80	$0.80	$0.80	$0.79
Muni Cable Car	$3.22	$3.22	$3.23	$3.22	$3.22	$3.20	$3.20	$3.19	$3.20	$3.54	$3.54	$3.53	$3.53	$3.53	$3.52	$3.51
BART	$3.09	$3.08	$3.09	$3.08	$3.08	$3.07	$3.06	$3.06	$3.06	$3.09	$3.08	$3.08	$3.08	$3.07	$3.07	$3.05
Caltrain	$4.06	$4.06	$4.07	$4.05	$4.05	$4.04	$4.03	$4.02	$4.03	$3.87	$3.86	$3.86	$3.86	$3.86	$3.85	$3.83
Weekday Service Miles																
Muni Bus	57,750	57,100	57,036	56,956	56,966	57,627	57,627	52,774	52,045	51,473	52,045	53,584	54,547	53,926	54,547	54,005
Muni Cable Car	1,352	1,348	1,337	1,334	1,334	1,338	1,338	1,336	1,340	1,336	1,340	1,336	1,340	1,336	1,340	1,232
Muni Rail	16,710	16,396	16,465	16,449	16,546	16,779	16,779	13,795	13,276	12,975	13,276	15,925	16,704	16,393	16,704	12,799
BART	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,170	27,166	27,166	27,166	27,166
Weekday Service Miles-Extrapolated																
Muni Bus	57,750	57,100	57,036	56,956	56,966	57,627	57,627	52,774	52,045	51,473	52,045	53,584	54,547	53,926	54,547	54,005
Muni Cable Car	1,352	1,348	1,337	1,334	1,334	1,338	1,338	1,336	1,340	1,336	1,340	1,336	1,340	1,336	1,340	1,232
Muni Rail	16,710	16,396	16,465	16,449	16,546	16,779	16,779	13,795	13,276	12,975	13,276	15,925	16,704	16,393	16,704	12,799
BART	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,176	27,170	27,166	27,166	27,166	27,166

Multi-Modal Performance Data

	Feb-2011	Mar-2011	Apr-2011	May-2011	Jun-2011	Jul-2011	Aug-2011	Sep-2011	Oct-2011	Nov-2011	Dec-2011	Jan-2012	Feb-2012	Mar-2012	Apr-2012	May-2012
Monthly Service Miles																
Muni Bus	1,477,417	1,477,417	1,477,417	1,477,417	1,477,417	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167	1,515,167
Muni Cable Car	23,917	23,917	23,917	23,917	23,917	25,250	25,250	25,250	25,250	25,250	25,250	25,250	25,250	25,250	25,250	25,250
Muni Rail	486,500	486,500	486,500	486,500	486,500	485,000	485,000	485,000	485,000	485,000	485,000	485,000	485,000	485,000	485,000	485,000
BART	5,278,917	5,278,917	5,278,917	5,278,917	5,278,917	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583	5,286,583
Caltrain	541,833	541,833	541,833	541,833	541,833	534,917	534,917	534,917	534,917	534,917	534,917	534,917	534,917	534,917	534,917	534,917
Average Weekday Ridership																
Muni Bus	491,906	491,906	491,906	491,906	491,906	510,151	510,151	510,151	510,151	510,151	510,151	510,151	510,151	510,151	510,151	510,151
Muni Cable Car	19,893	19,893	19,893	19,893	19,893	20,162	20,162	20,162	20,162	20,162	20,162	20,162	20,162	20,162	20,162	20,162
Muni Rail	161,398	161,398	161,398	161,398	161,398	163,980	163,980	163,980	163,980	163,980	163,980	163,980	163,980	163,980	163,980	163,980
BART	367,505	367,505	367,505	367,505	367,505	391,777	391,777	391,777	391,777	391,777	391,777	391,777	391,777	391,777	391,777	391,777
Caltrain	39,090	39,090	39,090	39,090	39,090	44,212	44,212	44,212	44,212	44,212	44,212	44,212	44,212	44,212	44,212	44,212
Average Weekday Ridership																
Muni Bus	493,509	493,512	516,458	521,283	499,697	492,836	512,709	531,730	515,078	491,879	465,621	470,854	499,611	479,422	500,787	489,294
BART	346,196	343,971	352,160	352,052	355,436	357,694	358,951	374,132	373,594	367,591	344,647	357,271	373,226	368,056	373,973	376,655
Cash Fare (2010$)																
Muni Bus+Rail	$1.97	$1.95	$1.94	$1.93	$1.93	$1.93	$1.93	$1.92	$1.93	$1.93	$1.93	$1.92	$1.92	$1.90	$1.90	$1.90
Muni Cable Car	$4.93	$4.88	$4.85	$4.83	$4.83	$5.79	$5.78	$5.77	$5.78	$5.78	$5.80	$5.77	$5.75	$5.70	$5.69	$5.69
BART	$3.49	$3.45	$3.43	$3.42	$3.42	$3.42	$3.41	$3.40	$3.41	$3.41	$3.42	$3.41	$3.39	$3.37	$3.35	$3.36
Average Fare (2010$)																
Muni Bus+Rail	$0.79	$0.78	$0.78	$0.77	$0.77	$0.78	$0.78	$0.77	$0.78	$0.78	$0.78	$0.77	$0.77	$0.77	$0.76	$0.76
Muni Cable Car	$3.49	$3.45	$3.43	$3.42	$3.42	$3.71	$3.70	$3.69	$3.70	$3.70	$3.71	$3.70	$3.68	$3.65	$3.64	$3.64
BART	$3.04	$3.01	$2.99	$2.98	$2.98	$2.97	$2.97	$2.96	$2.97	$2.97	$2.98	$2.96	$2.95	$2.93	$2.92	$2.92
Caltrain	$3.81	$3.77	$3.75	$3.73	$3.74	$4.09	$4.08	$4.07	$4.08	$4.08	$4.09	$4.08	$4.06	$4.03	$4.02	$4.02
Weekday Service Miles																
Muni Bus	54,467	55,167	55,167	54,533	55,329	54,746	55,422	54,777	55,463	54,841	55,501	54,991	55,275	55,975	55,975	55,336
Muni Cable Car	975	975	975	975	975	975	975	975	1,166	1,336	1,340	1,336	1,336	1,340	1,340	1,336
Muni Rail	12,881	13,060	13,060	12,821	13,060	12,810	13,060	12,821	12,990	12,735	12,927	12,807	12,802	12,927	12,927	12,743
BART	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166
Weekday Service Miles-Extrapolated																
Muni Bus	54,467	55,167	55,167	54,533	55,329	54,746	55,422	54,777	55,463	54,841	55,501	54,991	55,275	55,975	55,975	55,336
Muni Cable Car	975	975	975	975	975	975	975	975	1,166	1,336	1,340	1,336	1,336	1,340	1,340	1,336
Muni Rail	12,881	13,060	13,060	12,821	13,060	12,810	13,060	12,821	12,990	12,735	12,927	12,807	12,802	12,927	12,927	12,743
BART	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166	27,166

Multi-Modal Performance Data

	Jun-2012	Jul-2012	Aug-2012	Sep-2012	Oct-2012	Nov-2012	Dec-2012	Jan-2013
Monthly Service Miles								
Muni Bus	1,515,167	1,507,250	1,507,250	1,507,250	1,507,250	1,507,250	1,507,250	1,507,250
Muni Cable Car	25,250	25,000	25,000	25,000	25,000	25,000	25,000	25,000
Muni Rail	485,000	488,250	488,250	488,250	488,250	488,250	488,250	488,250
BART	5,286,583	5,471,000	5,471,000	5,471,000	5,471,000	5,471,000	5,471,000	5,471,000
Caltrain	534,917	549,250	549,250	549,250	549,250	549,250	549,250	549,250
Average Weekday Ridership								
Muni Bus	510,151	512,470	512,470	512,470	512,470	512,470	512,470	512,470
Muni Cable Car	20,162	18,962	18,962	18,962	18,962	18,962	18,962	18,962
Muni Rail	163,980	145,573	145,573	145,573	145,573	145,573	145,573	145,573
BART	391,777	420,396	420,396	420,396	420,396	420,396	420,396	420,396
Caltrain	44,212	49,031	49,031	49,031	49,031	49,031	49,031	49,031
Average Weekday Ridership								
Muni Bus	486,691	494,746	515,666	525,409	520,808	488,054	502,199	474,027
BART	373,047	375,689	384,295	402,955	416,932	393,328	375,092	373,945
Cash Fare (2010$)								
Muni Bus+Rail	$1.90	$1.90	$1.89	$1.88	$1.89	$1.89	$1.90	$1.89
Muni Cable Car	$5.70	$5.71	$5.68	$5.65	$5.66	$5.68	$5.70	$5.68
BART	$3.36	$3.42	$3.40	$3.38	$3.38	$3.40	$3.41	$3.40
Average Fare (2010$)								
Muni Bus+Rail	$0.77	$0.85	$0.84	$0.84	$0.84	$0.84	$0.84	$0.84
Muni Cable Car	$3.65	$3.73	$3.71	$3.69	$3.69	$3.71	$3.72	$3.71
BART	$2.93	$3.05	$3.04	$3.02	$3.02	$3.04	$3.05	$3.04
Caltrain	$4.03	$4.20	$4.17	$4.15	$4.16	$4.18	$4.19	$4.18
Weekday Service Miles								
Muni Bus	56,282	55,719	56,405	55,650	56,203	55,521	55,489	54,909
Muni Cable Car	1,340	1,336	1,340	1,336	1,340	1,336	1,336	1,333
Muni Rail	12,930	12,740	12,931	12,721	12,626	12,428	12,418	12,316
BART	27,166	27,102	27,102	27,315	27,372	27,372	27,372	27,372
Weekday Service Miles-Extrapolated								
Muni Bus	56,282	55,719	56,405	55,650	56,203	55,521	55,489	54,909
Muni Cable Car	1,340	1,336	1,340	1,336	1,340	1,336	1,336	1,333
Muni Rail	12,930	12,740	12,931	12,721	12,626	12,428	12,418	12,316
BART	27,166	27,102	27,102	27,315	27,372	27,372	27,372	27,372

	Feb-2013	Mar-2013	Apr-2013	May-2013	Jun-2013	Jul-2013	Aug-2013	Sep-2013
Monthly Service Miles								
Muni Bus	1,507,250	1,507,250	1,507,250	1,507,250	1,507,250	1,499,333	1,499,333	1,499,333
Muni Cable Car	25,000	25,000	25,000	25,000	25,000	24,833	24,833	24,833
Muni Rail	488,250	488,250	488,250	488,250	488,250	488,750	488,750	488,750
BART	5,471,000	5,471,000	5,471,000	5,471,000	5,471,000	5,563,583	5,563,583	5,563,583
Caltrain	549,250	549,250	549,250	549,250	549,250	563,500	563,500	563,500
Average Weekday Ridership								
Muni Bus	512,470	512,470	512,470	512,470	512,470	521,592	521,592	521,592
Muni Cable Car	18,962	18,962	18,962	18,962	18,962	19,299	19,299	19,299
Muni Rail	145,573	145,573	145,573	145,573	145,573	154,517	154,517	154,517
BART	420,396	420,396	420,396	420,396	420,396	432,913	432,913	432,913
Caltrain	49,031	49,031	49,031	49,031	49,031	53,466	53,466	53,466
Average Weekday Ridership								
Muni Bus	500,723	494,482	517,413	512,374	479,893	451,808	476,756	512,387
BART	394,692	394,380	400,435	397,396	396,414	321,382	399,712	415,803
Cash Fare (2010$)								
Muni Bus+Rail	$1.88	$1.87	$1.88	$1.87	$1.87	$1.87	$1.86	$1.86
Muni Cable Car	$5.64	$5.62	$5.63	$5.62	$5.60	$5.60	$5.59	$5.59
BART	$3.37	$3.36	$3.37	$3.36	$3.35	$3.35	$3.35	$3.34
Average Fare (2010$)								
Muni Bus+Rail	$0.84	$0.83	$0.83	$0.83	$0.83	$0.78	$0.78	$0.78
Muni Cable Car	$3.68	$3.67	$3.67	$3.67	$3.66	$3.47	$3.46	$3.46
BART	$3.01	$3.00	$3.01	$3.00	$3.00	$2.86	$2.86	$2.86
Caltrain	$4.14	$4.13	$4.13	$4.13	$4.12	$4.10	$4.10	$4.09
Weekday Service Miles								
Muni Bus	55,465	56,203	56,030	55,381	56,030	54,039	54,615	53,985
Muni Cable Car	1,336	1,340	1,340	1,336	1,340	1,336	1,340	1,336
Muni Rail	12,488	12,626	12,626	12,437	12,626	12,446	12,636	12,428
BART	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372
Weekday Service Miles-Extrapolated								
Muni Bus	55,465	56,203	56,030	55,381	56,030	54,039	54,615	53,985
Muni Cable Car	1,336	1,340	1,340	1,336	1,340	1,336	1,340	1,336
Muni Rail	12,488	12,626	12,626	12,437	12,626	12,446	12,636	12,428
BART	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372

Multi-Modal Performance Data

	Oct-2013	Nov-2013	Dec-2013	Jan-2014	Feb-2014	Mar-2014	Apr-2014	May-2014	Jun-2014	Jul-2014	Aug-2014	Sep-2014	Oct-2014	Nov-2014	Dec-2014	Jan-2015
Monthly Service Miles																
Muni Bus	1,499,333	1,499,333	1,499,333	1,499,333	1,499,333	1,499,333	1,499,333	1,499,333	1,499,333							
Muni Cable Car	24,833	24,833	24,833	24,833	24,833	24,833	24,833	24,833	24,833							
Muni Rail	488,750	488,750	488,750	488,750	488,750	488,750	488,750	488,750	488,750							
BART	5,563,583	5,563,583	5,563,583	5,563,583	5,563,583	5,563,583	5,563,583	5,563,583	5,563,583							
Caltrain	563,500	563,500	563,500	563,500	563,500	563,500	563,500	563,500	563,500							
Average Weekday Ridership																
Muni Bus	521,592	521,592	521,592	521,592	521,592	521,592	521,592	521,592	521,592							
Muni Cable Car	19,299	19,299	19,299	19,299	19,299	19,299	19,299	19,299	19,299							
Muni Rail	154,517	154,517	154,517	154,517	154,517	154,517	154,517	154,517	154,517							
BART	432,913	432,913	432,913	432,913	432,913	432,913	432,913	432,913	432,913							
Caltrain	53,466	53,466	53,466	53,466	53,466	53,466	53,466	53,466	53,466							
Average Weekday Ridership																
Muni Bus	503,591	506,678														
BART	359,904	394,897	377,892	381,090	398,669	403,177	403,898	409,927	410,312	414,896	419,255	430,163	441,428	424,763	400,071	410,847
Cash Fare (2010$)																
Muni Bus+Rail	$1.87	$1.87	$1.87	$1.86	$1.86	$1.85	$1.84	$1.83	$1.83	$1.83	$1.83	$2.06	$2.07	$2.08	$2.09	$2.10
Muni Cable Car	$5.60	$5.61	$5.61	$5.59	$5.57	$5.54	$5.52	$5.50	$5.49	$5.49	$5.50	$5.50	$5.51	$5.54	$5.57	$5.60
BART	$3.35	$3.36	$3.36	$3.52	$3.51	$3.49	$3.48	$3.46	$3.46	$3.46	$3.47	$3.46	$3.47	$3.49	$3.51	$3.53
Average Fare (2010$)																
Muni Bus+Rail	$0.78	$0.78	$0.78	$0.78	$0.77	$0.77	$0.77	$0.76	$0.76	$0.76	$0.76	$0.86	$0.86	$0.87	$0.87	$0.88
Muni Cable Car	$3.47	$3.47	$3.47	$3.46	$3.45	$3.43	$3.42	$3.40	$3.40	$3.40	$3.40	$3.40	$3.41	$3.43	$3.45	$3.46
BART	$2.87	$2.87	$2.87	$3.01	$3.00	$2.98	$2.97	$2.96	$2.96	$2.96	$2.96	$2.96	$2.97	$2.98	$3.00	$3.02
Caltrain	$4.10	$4.11	$4.11	$4.10	$4.08	$4.06	$4.04	$4.03	$4.02							
Weekday Service Miles																
Muni Bus	54,914	55,615	55,648	55,049												
Muni Cable Car	1,340	1,336	1,336	1,333												
Muni Rail	12,671	12,621	12,631	12,511												
BART	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	28,048	30,584	30,584
Weekday Service Miles-Extrapolated																
Muni Bus	54,914	55,615	55,648	55,049	55,049	55,049	55,049	55,049	55,049							
Muni Cable Car	1,340	1,336	1,336	1,333	1,333	1,333	1,333	1,333	1,333							
Muni Rail	12,671	12,621	12,631	12,511	12,511	12,511	12,511	12,511	12,511							
BART	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	27,372	28,048	30,584	30,584

Multi-Modal Performance Data

	Feb-2015	Mar-2015	Apr-2015	May-2015	Jun-2015	Jul-2015	Aug-2015	Sep-2015
Monthly Service Miles								
Muni Bus								
Muni Cable Car								
Muni Rail								
BART								
Caltrain								
Average Weekday Ridership								
Muni Bus								
Muni Cable Car								
Muni Rail								
BART								
Caltrain								
Average Weekday Ridership								
Muni Bus								
BART	420,906	423,526	429,910	424,312	432,869	427,467	429,750	445,103
Cash Fare (2010$)								
Muni Bus+Rail	$2.09	$2.08	$2.07	$2.06	$2.06	$2.06		
Muni Cable Car	$5.57	$5.54	$5.53	$5.50	$5.48	$6.40		
BART	$3.51	$3.49	$3.48	$3.47	$3.45	$3.45		
Average Fare (2010$)								
Muni Bus+Rail	$0.87	$0.87	$0.86	$0.86	$0.86	$0.86		
Muni Cable Car	$3.45	$3.43	$3.42	$3.41	$3.39	$3.96		
BART	$3.00	$2.98	$2.98	$2.96	$2.95	$2.95		
Caltrain								
Weekday Service Miles								
Muni Bus								
Muni Cable Car								
Muni Rail								
BART	30,584	30,584	30,584	30,584	30,584	30,584	30,584	30,770
Weekday Service Miles-Extrapolated								
Muni Bus								
Muni Cable Car								
Muni Rail								
BART	30,584	30,584	30,584	30,584	30,584	30,584	30,584	30,770

Drivers of Demand Data

	Source	Temporal Res	Geog Res	Trend	Jan-2000	Feb-2000	Mar-2000	Apr-2000	May-2000	Jun-2000	Jul-2000	Aug-2000
Population & Households												
Population	Census PopEst	Annual	County		776,395	776,643	776,892	777,140	777,388	777,637	777,885	778,133
Households	ACS	Annual	County					329,700	329,584	329,468	329,352	329,236
Housing Units	ACS	Annual	County					346,527	346,661	346,795	346,929	347,063
Housing Units	Planning Dept/Census	Date	Block									
Households, Income $0-15k	ACS	Annual	County					48,655	48,720	48,785	48,850	48,914
Households, Income $15-50k	ACS	Annual	County					101,522	101,325	101,127	100,930	100,732
Households, Income $50-100k	ACS	Annual	County					98,266	98,151	98,037	97,922	97,808
Households, Income $100k+	ACS	Annual	County					81,407	81,536	81,665	81,794	81,923
Households, 0 Vehicles	ACS	Annual	County					93,655	93,770	93,885	93,999	94,114
Median Household Income (2010$)	ACS	Annual	County					$80,889	$80,673	$80,130	$79,824	$79,703
Workers (at home location)												
Workers	LODES RAC/QCEW	Annual/Monthly	Block									
Workers, earning $0-15k	LODES RAC/QCEW	Annual/Monthly	Block									
Workers, earning $15-40k	LODES RAC/QCEW	Annual/Monthly	Block									
Workers, earning $40k+	LODES RAC/QCEW	Annual/Monthly	Block									
Employment (at work location)												
Total Employment	LODES WAC/QCEW	Monthly	Block		591,260	597,432	607,453	607,344	613,887	621,242	610,808	613,235
Retail Employment	LODES WAC/QCEW	Monthly	Block		46,460	45,809	45,714	45,460	46,050	46,642	47,542	47,719
Education and Health Employment	LODES WAC/QCEW	Monthly	Block		69,558	70,990	71,507	71,160	71,577	70,916	67,396	66,840
Leisure Employment	LODES WAC/QCEW	Monthly	Block		69,985	71,387	73,320	73,347	74,593	75,490	74,483	75,284
Other Employment	LODES WAC/QCEW	Monthly	Block		405,257	409,246	416,912	417,377	421,667	428,194	421,387	423,392
Employees, earning $0-15k	LODES WAC/QCEW	Monthly	Block									
Employees, earning $15-40k	LODES WAC/QCEW	Monthly	Block									
Employees, earning $40k+	LODES WAC/QCEW	Monthly	Block									
Average monthly earnings (2010$)	QCEW	Monthly	County		$6,071	$5,820	$5,558	$5,341	$5,357	$5,351	$5,361	$5,519
Jobs-Housing Balance												
Employees per Housing Unit	QCEW/Planning Dept	Monthly	Block									
Employees per Worker	LODES OD/QCEW	Annual/Monthly	Block									
Workers: Live & Work in SF	LODES OD/QCEW	Annual/Monthly	Block									
Workers: Live elswhere & work in SF	LODES OD/QCEW	Annual/Monthly	Block									
Workers: Live in SF & work elsewhere	LODES OD/QCEW	Annual/Monthly	Block									
Costs												
Average Fuel Price (2010$)	EIA	Monthly	MSA							$2.30	$2.40	$2.40
Average Fleet Efficiency (mpg)	BTS	Annual	US							21.86	21.90	21.92
Average Fuel Cost (2010$ / mi)	BTS/EIA	Annual/Monthly	US/MSA							$0.11	$0.11	$0.11
Average Auto Operating Cost (2010$/mile)	IRS	Annual	US							$0.13	$0.13	$0.13
Median Daily CBD Parking Cost (2010$)	Colliers	Annual	CBD									
Median Monthly CBD Parking Cost (2010$)	Colliers	Annual	CBD									
Bay Bridge Toll, Peak (2010$)	BATA	Monthly	Bridge		$2.58	$2.57	$2.55	$2.55	$2.54	$2.53	$2.52	$2.52
Bay Bridge Toll, Off-Peak (2010$)	BATA	Monthly	Bridge		$2.58	$2.57	$2.55	$2.55	$2.54	$2.53	$2.52	$2.52
Bay Bridge Toll, Carpools (2010$)	BATA	Monthly	Bridge		$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	BATA	Monthly	Bridge		$3.45	$3.43	$3.40	$3.40	$3.39	$3.38	$3.37	$3.37
Golden Gate Bridge Toll, Carpools (2010$)	BATA	Monthly	Bridge		$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	BLS	Monthly	US City Avg		169	170	171	171	172	172	173	173

Drivers of Demand Data

	Sep-2000	Oct-2000	Nov-2000	Dec-2000	Jan-2001	Feb-2001	Mar-2001	Apr-2001	May-2001	Jun-2001	Jul-2001	Aug-2001	Sep-2001	Oct-2001	Nov-2001	Dec-2001
Population & Households																
Population	778,382	778,630	778,878	779,127	779,375	779,623	779,872	780,120	780,368	780,617	780,865	780,187	779,508	778,830	778,151	777,473
Households	329,121	329,005	328,889	328,773	328,657	328,541	328,425	328,309	328,193	328,078	327,962	327,846	327,730	327,614	327,498	327,382
Housing Units	347,197	347,330	347,464	347,598	347,732	347,866	348,000	348,134	348,268	348,402	348,536	348,669	348,803	348,937	349,071	349,205
Housing Units					354,378	354,607	354,962	355,054	355,211	355,284	355,417	355,480	355,614	355,977	356,073	356,157
Households, Income $0-15k	48,979	49,044	49,109	49,174	49,239	49,304	49,368	49,433	49,498	49,563	49,628	49,693	49,758	49,822	49,887	49,952
Households, Income $15-50k	100,535	100,337	100,140	99,942	99,745	99,547	99,350	99,152	98,955	98,757	98,560	98,362	98,165	97,967	97,770	97,572
Households, Income $50-100k	97,693	97,578	97,464	97,349	97,235	97,120	97,005	96,891	96,776	96,662	96,547	96,432	96,318	96,203	96,089	95,974
Households, Income $100k+	82,052	82,181	82,310	82,439	82,568	82,697	82,825	82,954	83,083	83,212	83,341	83,470	83,599	83,728	83,857	83,986
Households, 0 Vehicles	94,229	94,344	94,458	94,573	94,688	94,803	94,918	95,032	95,147	95,262	95,377	95,491	95,606	95,721	95,836	95,951
Median Household Income (2010$)	$79,169	$78,912	$78,747	$78,672	$78,058	$77,628	$77,333	$76,909	$76,445	$76,198	$76,295	$76,177	$75,718	$75,855	$75,866	$76,048
Workers (at home location)																
Workers																
Workers, earning $0-15k																
Workers, earning $15-40k																
Workers, earning $40k+																
Employment (at work location)																
Total Employment	621,894	614,823	617,790	622,939	599,433	601,733	606,305	591,651	591,823	593,416	580,483	577,207	576,828	570,711	570,341	573,092
Retail Employment	47,965	47,991	49,616	51,069	46,873	45,946	46,063	45,702	45,813	46,282	45,567	45,582	45,500	44,346	45,471	46,664
Education and Health Employment	68,309	68,471	69,046	68,775	84,748	85,966	86,921	85,897	86,094	86,203	83,121	82,757	85,969	87,785	89,595	89,244
Leisure Employment	77,236	74,987	74,354	74,359	71,688	73,085	74,070	72,787	75,023	74,813	74,067	74,533	73,310	70,857	69,378	69,377
Other Employment	428,384	423,374	424,774	428,736	396,124	396,736	399,251	387,265	384,893	386,118	377,728	374,335	372,049	367,723	365,897	367,807
Employees, earning $0-15k																
Employees, earning $15-40k																
Employees, earning $40k+																
Average monthly earnings (2010$)	$5,647	$5,795	$6,216	$6,644	$7,024	$6,440	$5,871	$5,295	$5,254	$5,227	$5,224	$5,355	$5,461	$5,610	$5,810	$6,024
Jobs-Housing Balance																
Employees per Housing Unit					1.69	1.70	1.71	1.67	1.67	1.67	1.63	1.62	1.62	1.60	1.60	1.61
Employees per Worker																
Workers: Live & Work in SF																
Workers: Live elswhere & work in SF																
Workers: Live in SF & work elsewhere																
Costs																
Average Fuel Price (2010$)	$2.51	$2.50	$2.46	$2.37	$2.24	$2.23	$2.29	$2.44	$2.53	$2.52	$2.37	$2.14	$2.29	$2.19	$2.01	$1.74
Average Fleet Efficiency (mpg)	21.93	21.95	21.97	21.98	22.00	22.02	22.03	22.05	22.07	22.08	22.10	22.09	22.08	22.08	22.07	22.06
Average Fuel Cost (2010$ / mi)	$0.11	$0.11	$0.11	$0.11	$0.10	$0.10	$0.10	$0.11	$0.11	$0.11	$0.11	$0.10	$0.10	$0.10	$0.09	$0.08
Average Auto Operating Cost (2010$/mile)	$0.13	$0.13	$0.13	$0.13	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15	$0.15
Median Daily CBD Parking Cost (2010$)																
Median Monthly CBD Parking Cost (2010$)																
Bay Bridge Toll, Peak (2010$)	$2.51	$2.51	$2.50	$2.51	$2.49	$2.48	$2.48	$2.47	$2.45	$2.45	$2.46	$2.46	$2.45	$2.45	$2.46	$2.47
Bay Bridge Toll, Off-Peak (2010$)	$2.51	$2.51	$2.50	$2.51	$2.49	$2.48	$2.48	$2.47	$2.45	$2.45	$2.46	$2.46	$2.45	$2.45	$2.46	$2.47
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$3.35	$3.35	$3.34	$3.35	$3.33	$3.31	$3.30	$3.29	$3.28	$3.27	$3.69	$3.69	$3.67	$3.68	$3.69	$3.70
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	174	174	174	174	175	176	176	177	178	178	178	178	178	178	177	177

Drivers of Demand Data

	Jan-2002	Feb-2002	Mar-2002	Apr-2002	May-2002	Jun-2002	Jul-2002	Aug-2002	Sep-2002	Oct-2002	Nov-2002	Dec-2002	Jan-2003	Feb-2003	Mar-2003	Apr-2003
Population & Households																
Population	776,794	776,116	775,437	774,759	774,080	773,402	772,723	772,176	771,629	771,082	770,535	769,988	769,442	768,895	768,348	767,801
Households	327,266	327,150	327,035	326,919	326,803	326,687	326,571	326,455	326,339	326,223	326,107	325,992	325,876	325,760	325,644	325,528
Housing Units	349,339	349,473	349,607	349,741	349,875	350,009	350,142	350,276	350,410	350,544	350,678	350,812	350,946	351,080	351,214	351,348
Housing Units	356,197	356,443	356,517	356,599	356,805	356,845	357,492	357,710	358,007	358,203	358,449	358,691	358,777	359,132	359,185	359,997
Households, Income $0-15k	50,017	50,082	50,147	50,212	50,276	50,341	50,406	50,471	50,536	50,601	50,666	50,730	50,795	50,860	50,925	50,990
Households, Income $15-50k	97,375	97,178	96,980	96,783	96,585	96,388	96,190	95,993	95,795	95,598	95,400	95,203	95,005	94,808	94,610	94,413
Households, Income $50-100k	95,859	95,745	95,630	95,516	95,401	95,286	95,172	95,057	94,943	94,828	94,713	94,599	94,484	94,369	94,255	94,140
Households, Income $100k+	84,115	84,244	84,373	84,502	84,631	84,760	84,889	85,018	85,147	85,276	85,405	85,533	85,662	85,791	85,920	86,049
Households, 0 Vehicles	96,065	96,180	96,295	96,410	96,524	96,639	96,754	96,869	96,984	97,098	97,213	97,328	97,443	97,557	97,672	97,787
Median Household Income (2010$)	$75,758	$75,342	$74,803	$74,271	$74,154	$73,997	$73,798	$73,437	$73,200	$72,963	$72,848	$72,893	$72,457	$71,789	$71,246	$71,287
Workers (at home location)																
Workers	386,633	387,030	390,292	388,070	391,253	391,862	385,841	388,077	389,581	388,177	392,146	395,037	378,700	380,431	382,680	382,618
Workers, earning $0-15k	101,238	101,425	102,365	101,868	102,789	103,036	101,539	102,215	102,700	102,418	103,555	104,410	100,180	100,726	101,412	101,486
Workers, earning $15-40k	128,922	128,918	129,867	128,989	129,906	129,967	127,829	128,428	128,782	128,174	129,338	130,144	124,618	125,043	125,635	125,467
Workers, earning $40k+	156,473	156,686	158,060	157,214	158,557	158,859	156,472	157,434	158,099	157,585	159,253	160,484	153,903	154,662	155,633	155,665
Employment (at work location)																
Total Employment	546,351	546,365	550,416	546,727	550,647	550,937	541,909	544,480	546,015	543,469	548,439	551,888	528,490	530,324	532,872	532,194
Retail Employment	43,979	43,392	43,378	43,029	43,194	43,473	43,146	43,122	43,555	43,392	45,200	46,011	43,898	43,117	42,829	42,343
Education and Health Employment	87,889	88,704	89,343	87,870	89,034	88,132	84,607	86,599	88,108	89,414	90,359	89,785	86,653	88,262	88,922	89,525
Leisure Employment	65,541	65,499	67,199	69,009	71,327	71,209	69,740	71,233	70,649	70,386	70,061	69,498	67,511	67,571	68,323	69,058
Other Employment	348,942	348,770	350,496	346,819	347,092	348,123	344,416	343,526	343,703	340,277	342,819	346,594	330,428	331,374	332,798	331,268
Employees, earning $0-15k	129,970	130,076	131,145	130,371	131,412	131,589	129,539	130,262	130,739	130,239	131,541	132,482	126,975	127,526	128,251	128,202
Employees, earning $15-40k	179,068	178,860	179,970	178,547	179,607	179,480	176,319	176,932	177,206	176,153	177,533	178,416	170,625	170,989	171,578	171,126
Employees, earning $40k+	237,314	237,429	239,301	237,809	239,627	239,868	236,051	237,286	238,071	237,077	239,364	240,990	230,890	231,809	233,042	232,866
Average monthly earnings (2010$)	$6,201	$5,833	$5,459	$5,089	$5,107	$5,122	$5,134	$5,234	$5,343	$5,451	$5,610	$5,781	$5,914	$5,580	$5,259	$4,983
Jobs-Housing Balance																
Employees per Housing Unit	1.53	1.53	1.54	1.53	1.54	1.54	1.52	1.52	1.53	1.52	1.53	1.54	1.47	1.48	1.48	1.48
Employees per Worker	1.41	1.41	1.41	1.41	1.41	1.41	1.40	1.40	1.40	1.40	1.40	1.40	1.40	1.39	1.39	1.39
Workers: Live & Work in SF	241,159	241,185	242,993	241,384	243,135	243,283	239,317	240,473	241,171	240,067	242,284	243,829	233,512	234,343	235,490	235,212
Workers: Live elswhere & work in SF	303,524	303,528	305,775	303,723	305,897	306,055	301,036	302,461	303,310	301,892	304,650	306,562	293,561	294,576	295,988	295,608
Workers: Live in SF & work elsewhere	144,564	144,941	146,395	145,794	147,225	147,692	145,658	146,740	147,550	147,259	149,010	150,358	144,380	145,283	146,388	146,612
Costs																
Average Fuel Price (2010$)	$1.63	$1.61	$1.84	$2.07	$2.03	$2.06	$2.08	$2.07	$2.03	$1.97	$2.06	$2.04	$2.14	$2.32	$2.60	$2.53
Average Fleet Efficiency (mpg)	22.05	22.04	22.03	22.03	22.02	22.01	22.00	22.02	22.03	22.05	22.07	22.08	22.10	22.12	22.13	22.15
Average Fuel Cost (2010$ / mi)	$0.07	$0.07	$0.08	$0.09	$0.09	$0.09	$0.09	$0.09	$0.09	$0.09	$0.09	$0.09	$0.10	$0.11	$0.12	$0.11
Average Auto Operating Cost (2010$/mile)	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.14	$0.14	$0.14	$0.14
Median Daily CBD Parking Cost (2010$)																
Median Monthly CBD Parking Cost (2010$)																
Bay Bridge Toll, Peak (2010$)	$2.46	$2.45	$2.44	$2.43	$2.43	$2.42	$2.42	$2.41	$2.41	$2.41	$2.41	$2.41	$2.40	$2.38	$2.37	$2.37
Bay Bridge Toll, Off-Peak (2010$)	$2.46	$2.45	$2.44	$2.43	$2.43	$2.42	$2.42	$2.41	$2.41	$2.41	$2.41	$2.41	$2.40	$2.38	$2.37	$2.37
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$3.69	$3.68	$3.66	$3.64	$3.64	$3.64	$3.63	$3.62	$4.82	$4.81	$4.81	$4.82	$4.80	$4.76	$4.74	$4.75
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	177	178	179	180	180	180	180	181	181	181	181	181	182	183	184	184

Drivers of Demand Data

	May-2003	Jun-2003	Jul-2003	Aug-2003	Sep-2003	Oct-2003	Nov-2003	Dec-2003	Jan-2004	Feb-2004	Mar-2004	Apr-2004	May-2004	Jun-2004	Jul-2004	Aug-2004
Population & Households																
Population	767,254	766,707	766,160	765,757	765,354	764,951	764,548	764,145	763,743	763,340	762,937	762,534	762,131	761,728	761,325	761,466
Households	325,412	325,296	325,180	325,064	324,949	324,833	324,717	324,601	324,485	324,369	324,253	324,137	324,021	323,906	323,790	323,674
Housing Units	351,481	351,615	351,749	351,883	352,017	352,151	352,285	352,419	352,553	352,687	352,821	352,954	353,088	353,222	353,356	353,490
Housing Units	360,163	360,235	360,522	360,963	361,058	361,243	361,347	361,521	361,639	361,793	362,005	362,111	362,277	362,341	362,982	363,000
Households, Income $0-15k	51,055	51,120	51,184	51,249	51,314	51,379	51,444	51,509	51,574	51,638	51,703	51,768	51,833	51,898	51,963	52,028
Households, Income $15-50k	94,215	94,018	93,820	93,623	93,425	93,228	93,031	92,833	92,636	92,438	92,241	92,043	91,846	91,648	91,451	91,253
Households, Income $50-100k	94,026	93,911	93,796	93,682	93,567	93,453	93,338	93,223	93,109	92,994	92,880	92,765	92,650	92,536	92,421	92,307
Households, Income $100k+	86,178	86,307	86,436	86,565	86,694	86,823	86,952	87,081	87,210	87,339	87,468	87,597	87,726	87,855	87,984	88,113
Households, 0 Vehicles	97,902	98,017	98,131	98,246	98,361	98,476	98,590	98,705	98,820	98,935	99,050	99,164	99,279	99,394	99,509	99,623
Median Household Income (2010$)	$71,290	$71,098	$70,907	$70,525	$70,183	$70,146	$70,223	$70,185	$69,731	$69,244	$68,689	$68,359	$67,850	$67,525	$67,522	$67,375
Workers (at home location)																
Workers	383,680	384,997	381,267	383,305	384,354	384,600	385,345	387,965	368,172	370,826	370,900	370,704	372,218	372,067	373,340	372,847
Workers, earning $0-15k	101,859	102,302	101,403	101,657	101,646	101,422	101,328	101,724	96,257	96,670	96,409	96,077	96,187	95,865	95,909	95,566
Workers, earning $15-40k	125,666	125,947	124,577	125,080	125,258	125,174	125,252	125,938	119,356	120,057	119,922	119,699	120,028	119,819	120,068	119,894
Workers, earning $40k+	156,154	156,748	155,287	156,568	157,449	158,004	158,764	160,303	152,560	154,099	154,570	154,928	156,004	156,383	157,363	157,387
Employment (at work location)																
Total Employment	533,073	534,299	528,521	531,733	533,576	534,307	535,732	539,769	512,605	516,677	517,158	517,262	519,755	519,924	522,084	519,424
Retail Employment	42,520	42,772	42,781	43,207	43,113	43,413	45,555	46,437	42,613	42,264	42,062	42,267	42,491	42,674	42,870	43,055
Education and Health Employment	89,758	89,311	87,746	88,088	90,268	91,553	92,344	92,655	90,747	92,793	91,732	93,296	92,966	92,163	92,465	90,560
Leisure Employment	70,007	71,327	70,160	72,702	72,117	72,221	70,079	70,636	66,959	68,478	69,236	70,156	72,115	72,023	72,669	72,540
Other Employment	330,788	330,889	327,834	327,736	328,078	327,120	327,754	330,041	312,286	313,142	314,128	311,543	312,183	313,064	314,080	313,269
Employees, earning $0-15k	128,528	128,939	127,660	127,813	127,631	127,179	126,891	127,215	120,213	120,563	120,070	119,490	119,458	118,889	118,772	118,060
Employees, earning $15-40k	171,173	171,327	169,236	169,747	169,816	169,529	169,460	170,212	161,148	161,926	161,574	161,104	161,376	160,924	161,086	160,461
Employees, earning $40k+	233,373	234,033	231,625	234,173	236,129	237,598	239,381	242,341	231,244	234,188	235,513	236,668	238,921	240,112	242,226	240,903
Average monthly earnings (2010$)	$5,018	$5,039	$5,061	$5,215	$5,371	$5,549	$5,779	$5,999	$6,184	$5,821	$5,456	$5,113	$5,091	$5,082	$5,098	$5,353
Jobs-Housing Balance																
Employees per Housing Unit	1.48	1.48	1.47	1.47	1.48	1.48	1.48	1.49	1.42	1.43	1.43	1.43	1.43	1.43	1.44	1.43
Employees per Worker	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.39	1.40	1.40	1.40	1.40	1.39
Workers: Live & Work in SF	235,622	236,186	233,654	234,900	235,540	235,688	236,142	237,746	225,614	227,238	227,281	227,158	228,084	227,988	228,766	227,898
Workers: Live elswhere & work in SF	296,093	296,770	293,557	295,529	296,742	297,337	298,319	300,757	285,802	288,255	288,706	288,946	290,522	290,799	292,191	290,397
Workers: Live in SF & work elsewhere	147,268	148,026	146,844	147,627	148,028	148,120	148,405	149,412	141,787	142,806	142,832	142,754	143,335	143,275	143,762	144,126
Costs																
Average Fuel Price (2010$)	$2.32	$2.25	$2.25	$2.35	$2.50	$2.24	$2.10	$2.01	$2.06	$2.26	$2.54	$2.55	$2.67	$2.70	$2.60	$2.48
Average Fleet Efficiency (mpg)	22.17	22.18	22.20	22.23	22.25	22.28	22.30	22.33	22.35	22.38	22.40	22.43	22.45	22.48	22.50	22.47
Average Fuel Cost (2010$ / mi)	$0.10	$0.10	$0.10	$0.11	$0.11	$0.10	$0.09	$0.09	$0.09	$0.10	$0.11	$0.11	$0.12	$0.12	$0.12	$0.11
Average Auto Operating Cost (2010$/mile)	$0.14	$0.14	$0.14	$0.14	$0.14	$0.14	$0.14	$0.14	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16
Median Daily CBD Parking Cost (2010$)																
Median Monthly CBD Parking Cost (2010$)																
Bay Bridge Toll, Peak (2010$)	$2.38	$2.37	$2.37	$2.36	$2.35	$2.36	$2.36	$2.37	$2.35	$2.34	$2.33	$2.32	$2.31	$2.30	$3.45	$3.45
Bay Bridge Toll, Off-Peak (2010$)	$2.38	$2.37	$2.37	$2.36	$2.35	$2.36	$2.36	$2.37	$2.35	$2.34	$2.33	$2.32	$2.31	$2.30	$3.45	$3.45
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$4.75	$4.75	$4.74	$4.72	$4.71	$4.71	$4.73	$4.73	$4.71	$4.68	$4.65	$4.64	$4.61	$4.60	$4.61	$4.60
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	184	184	184	185	185	185	185	184	185	186	187	188	189	190	189	190

Drivers of Demand Data

	Sep-2004	Oct-2004	Nov-2004	Dec-2004	Jan-2005	Feb-2005	Mar-2005	Apr-2005	May-2005	Jun-2005	Jul-2005	Aug-2005	Sep-2005	Oct-2005	Nov-2005	Dec-2005
Population & Households																
Population	761,607	761,748	761,889	762,030	762,171	762,311	762,452	762,593	762,734	762,875	763,016	763,486	763,955	764,425	764,894	765,364
Households	323,558	323,442	323,326	323,210	323,094	322,978	322,863	322,747	322,631	322,515	322,399	322,411	322,424	322,436	322,448	322,461
Housing Units	353,624	353,758	353,892	354,026	354,160	354,293	354,427	354,561	354,695	354,829	354,963	355,090	355,217	355,344	355,471	355,598
Housing Units	363,080	363,362	363,484	363,505	363,532	363,688	363,844	364,000	364,156	364,312	364,468	364,624	364,780	364,936	365,092	365,248
Households, Income $0-15k	52,092	52,157	52,222	52,287	52,352	52,417	52,482	52,546	52,611	52,676	52,741	51,976	51,211	50,446	49,681	48,916
Households, Income $15-50k	91,056	90,858	90,661	90,463	90,266	90,068	89,871	89,673	89,476	89,278	89,081	88,743	88,405	88,066	87,728	87,390
Households, Income $50-100k	92,192	92,077	91,963	91,848	91,734	91,619	91,504	91,390	91,275	91,161	91,046	90,876	90,706	90,536	90,366	90,196
Households, Income $100k+	88,241	88,370	88,499	88,628	88,757	88,886	89,015	89,144	89,273	89,402	89,531	90,817	92,102	93,388	94,674	95,959
Households, 0 Vehicles	99,738	99,853	99,968	100,083	100,197	100,312	100,427	100,542	100,656	100,771	100,886	100,170	99,455	98,739	98,024	97,308
Median Household Income (2010$)	$67,123	$66,662	$66,517	$66,652	$66,403	$65,913	$65,293	$64,749	$64,708	$64,567	$64,162	$64,576	$64,528	$65,128	$66,391	$67,400
Workers (at home location)																
Workers	376,124	378,243	381,574	386,022	374,756	379,479	382,899	385,909	389,535	390,956	393,095	390,711	394,812	395,887	399,296	400,471
Workers, earning $0-15k	96,189	96,514	97,147	98,061	94,988	95,973	96,624	97,170	97,869	98,012	98,334	97,374	98,031	97,934	98,412	98,337
Workers, earning $15-40k	120,933	121,598	122,654	124,068	120,432	121,935	123,018	123,970	125,119	125,561	126,232	124,825	125,490	125,188	125,620	125,345
Workers, earning $40k+	159,003	160,130	161,773	163,893	159,336	161,572	163,257	164,769	166,547	167,383	168,529	168,512	171,291	172,765	175,264	176,788
Employment (at work location)																
Total Employment	522,013	522,976	525,599	529,731	512,346	516,865	519,578	521,713	524,658	524,619	525,537	522,698	528,536	530,324	535,241	537,166
Retail Employment	42,879	43,225	44,590	46,166	43,044	42,447	42,346	42,337	42,432	42,642	42,736	43,125	43,354	43,516	44,987	46,256
Education and Health Employment	92,746	92,526	93,388	93,924	91,079	93,407	94,100	94,612	95,041	92,670	93,769	88,990	93,814	94,931	96,040	95,980
Leisure Employment	73,702	72,528	71,610	72,375	68,629	69,037	69,596	71,573	72,617	73,006	72,702	73,098	73,849	73,598	74,142	73,228
Other Employment	312,686	314,697	316,011	317,266	309,594	311,974	313,536	313,191	314,568	316,301	316,330	317,485	317,519	318,279	320,072	321,702
Employees, earning $0-15k	118,541	118,651	119,137	119,964	115,920	116,835	117,340	117,713	118,268	118,149	118,245	117,081	117,861	117,735	118,299	118,199
Employees, earning $15-40k	161,459	161,955	162,967	164,450	159,248	160,850	161,894	162,759	163,879	164,069	164,559	162,665	163,472	163,018	163,521	163,103
Employees, earning $40k+	242,014	242,370	243,495	245,317	237,177	239,179	240,344	241,241	242,511	242,401	242,733	242,952	247,203	249,571	253,421	255,864
Average monthly earnings (2010$)	$5,599	$5,825	$5,970	$6,140	$6,275	$5,922	$5,562	$5,213	$5,290	$5,359	$5,406	$5,622	$5,796	$6,025	$6,280	$6,512
Jobs-Housing Balance																
Employees per Housing Unit	1.44	1.44	1.45	1.46	1.41	1.42	1.43	1.43	1.44	1.44	1.44	1.43	1.45	1.45	1.47	1.47
Employees per Worker	1.39	1.38	1.38	1.37	1.37	1.36	1.36	1.35	1.35	1.34	1.34	1.34	1.34	1.34	1.34	1.34
Workers: Live & Work in SF	229,334	230,058	231,515	233,642	226,271	228,566	230,068	231,317	232,929	233,218	233,934	232,575	235,078	235,778	237,869	238,630
Workers: Live elswhere & work in SF	291,536	291,764	292,916	294,905	284,922	287,127	288,324	289,196	290,515	290,178	290,370	288,906	292,238	293,331	296,156	297,326
Workers: Live in SF & work elsewhere	145,948	147,325	149,179	151,478	147,598	150,003	151,901	153,642	155,635	156,751	158,157	157,134	158,719	159,087	160,393	160,801
Costs																
Average Fuel Price (2010$)	$2.46	$2.75	$2.74	$2.49	$2.27	$2.40	$2.63	$2.98	$2.91	$2.73	$2.92	$3.07	$3.36	$3.24	$2.81	$2.50
Average Fleet Efficiency (mpg)	22.43	22.40	22.37	22.33	22.30	22.27	22.23	22.20	22.17	22.13	22.10	22.13	22.17	22.20	22.23	22.27
Average Fuel Cost (2010$ / mi)	$0.11	$0.12	$0.12	$0.11	$0.10	$0.11	$0.12	$0.13	$0.13	$0.12	$0.13	$0.14	$0.15	$0.15	$0.13	$0.11
Average Auto Operating Cost (2010$/mile)	$0.16	$0.16	$0.16	$0.16	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17
Median Daily CBD Parking Cost (2010$)																
Median Monthly CBD Parking Cost (2010$)																
Bay Bridge Toll, Peak (2010$)	$3.44	$3.43	$3.42	$3.44	$3.43	$3.41	$3.38	$3.36	$3.37	$3.36	$3.35	$3.33	$3.29	$3.28	$3.31	$3.32
Bay Bridge Toll, Off-Peak (2010$)	$3.44	$3.43	$3.42	$3.44	$3.43	$3.41	$3.38	$3.36	$3.37	$3.36	$3.35	$3.33	$3.29	$3.28	$3.31	$3.32
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$4.59	$4.57	$4.57	$4.58	$4.57	$4.55	$4.51	$4.48	$4.49	$4.48	$4.46	$4.44	$4.39	$4.38	$4.41	$4.43
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	190	191	191	190	191	192	193	195	194	195	195	196	199	199	198	197

Drivers of Demand Data

	Jan-2006	Feb-2006	Mar-2006	Apr-2006	May-2006	Jun-2006	Jul-2006	Aug-2006	Sep-2006	Oct-2006	Nov-2006	Dec-2006	Jan-2007	Feb-2007	Mar-2007	Apr-2007
Population & Households																
Population	765,833	766,303	766,772	767,242	767,711	768,181	768,650	769,485	770,321	771,156	771,991	772,826	773,662	774,497	775,332	776,167
Households	322,473	322,485	322,498	322,510	322,522	322,535	322,547	322,497	322,447	322,397	322,347	322,297	322,247	322,197	322,147	322,097
Housing Units	355,725	355,851	355,978	356,105	356,232	356,359	356,486	356,598	356,711	356,823	356,935	357,047	357,160	357,272	357,384	357,496
Housing Units	365,404	365,642	365,705	365,738	365,796	365,854	366,169	366,323	366,537	366,688	366,959	367,328	367,432	367,642	367,711	367,847
Households, Income $0-15k	48,151	47,385	46,620	45,855	45,090	44,325	43,560	43,281	43,002	42,723	42,444	42,165	41,886	41,607	41,328	41,049
Households, Income $15-50k	87,052	86,713	86,375	86,037	85,699	85,360	85,022	84,993	84,965	84,936	84,908	84,879	84,851	84,822	84,793	84,765
Households, Income $50-100k	90,026	89,856	89,686	89,516	89,346	89,176	89,006	88,697	88,389	88,080	87,772	87,463	87,155	86,846	86,537	86,229
Households, Income $100k+	97,245	98,531	99,816	101,102	102,388	103,673	104,959	105,525	106,091	106,658	107,224	107,790	108,356	108,922	109,488	110,055
Households, 0 Vehicles	96,593	95,877	95,161	94,446	93,730	93,015	92,299	92,533	92,768	93,002	93,237	93,471	93,706	93,940	94,174	94,409
Median Household Income (2010$)	$67,623	$68,219	$68,571	$68,714	$69,093	$69,673	$70,182	$70,269	$70,842	$71,455	$71,790	$71,910	$71,918	$71,761	$71,337	$71,099
Workers (at home location)																
Workers	388,832	392,839	394,854	394,994	398,246	399,999	399,121	395,065	401,352	404,039	405,076	405,587	400,529	403,554	405,023	405,200
Workers, earning $0-15k	95,127	95,754	95,891	95,574	96,007	96,077	95,517	94,070	95,088	95,247	95,016	94,663	93,021	93,262	93,142	92,726
Workers, earning $15-40k	121,080	121,703	121,702	121,123	121,496	121,408	120,523	119,049	120,694	121,252	121,314	121,221	119,468	120,128	120,326	120,140
Workers, earning $40k+	172,625	175,382	177,261	178,297	180,742	182,514	183,082	181,946	185,571	187,541	188,746	189,703	188,041	190,164	191,556	192,334
Employment (at work location)																
Total Employment	521,893	527,610	530,656	531,182	535,894	538,592	537,746	532,442	541,078	544,862	546,421	547,270	540,602	544,841	546,980	547,373
Retail Employment	42,687	41,912	41,712	41,950	42,189	42,416	42,421	42,604	43,067	43,721	45,230	46,049	44,633	43,899	43,581	43,157
Education and Health Employment	92,794	95,519	96,046	96,555	96,738	96,524	95,311	90,510	97,120	97,450	98,082	98,240	95,806	98,475	98,813	98,340
Leisure Employment	70,365	71,705	72,159	73,593	75,607	75,964	75,307	75,120	76,066	77,204	76,075	75,585	72,542	73,288	74,027	75,359
Other Employment	316,047	318,474	320,739	319,084	321,360	323,688	324,707	324,208	324,825	326,487	327,034	327,396	327,621	329,179	330,559	330,517
Employees, earning $0-15k	114,331	115,075	115,230	114,838	115,350	115,424	114,740	113,344	114,916	115,454	115,521	115,438	113,775	114,411	114,605	114,435
Employees, earning $15-40k	157,495	158,245	158,184	157,372	157,797	157,622	156,413	154,293	156,216	156,729	156,602	156,274	153,811	154,458	154,509	154,068
Employees, earning $40k+	250,067	254,291	257,242	258,972	262,747	265,546	266,593	264,804	269,946	272,679	274,299	275,558	273,016	275,972	277,866	278,870
Average monthly earnings (2010$)	$6,668	$6,227	$5,768	$5,298	$5,301	$5,319	$5,332	$5,627	$5,961	$6,302	$6,577	$6,832	$7,075	$6,582	$6,071	$5,583
Jobs-Housing Balance																
Employees per Housing Unit	1.43	1.44	1.45	1.45	1.47	1.47	1.47	1.45	1.48	1.49	1.49	1.49	1.47	1.48	1.49	1.49
Employees per Worker	1.34	1.34	1.34	1.34	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35	1.35
Workers: Live & Work in SF	231,753	234,200	235,460	235,603	237,601	238,706	238,240	236,023	239,986	241,798	242,623	243,132	240,298	242,312	243,391	243,694
Workers: Live elswhere & work in SF	288,973	292,240	294,029	294,421	297,134	298,731	298,362	295,252	299,873	301,803	302,500	302,805	298,953	301,136	302,157	302,215
Workers: Live in SF & work elsewhere	156,066	157,612	158,358	158,353	159,594	160,235	159,822	157,966	160,246	161,086	161,268	161,241	159,005	159,981	160,339	160,187
Costs																
Average Fuel Price (2010$)	$2.64	$2.76	$2.84	$3.18	$3.61	$3.48	$3.46	$3.39	$3.13	$2.78	$2.72	$2.82	$2.91	$3.03	$3.41	$3.62
Average Fleet Efficiency (mpg)	22.30	22.33	22.37	22.40	22.43	22.47	22.50	22.53	22.57	22.60	22.63	22.67	22.70	22.73	22.77	22.80
Average Fuel Cost (2010$ / mi)	$0.12	$0.12	$0.13	$0.14	$0.16	$0.16	$0.15	$0.15	$0.14	$0.12	$0.12	$0.12	$0.13	$0.13	$0.15	$0.16
Average Auto Operating Cost (2010$/mile)	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.22	$0.22	$0.22	$0.22
Median Daily CBD Parking Cost (2010$)																
Median Monthly CBD Parking Cost (2010$)																
Bay Bridge Toll, Peak (2010$)	$3.30	$3.29	$3.27	$3.25	$3.23	$3.22	$3.21	$3.21	$3.22	$3.24	$3.25	$3.24	$4.31	$4.29	$4.25	$4.22
Bay Bridge Toll, Off-Peak (2010$)	$3.30	$3.29	$3.27	$3.25	$3.23	$3.22	$3.21	$3.21	$3.22	$3.24	$3.25	$3.24	$4.31	$4.29	$4.25	$4.22
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$4.40	$4.39	$4.37	$4.33	$4.31	$4.30	$4.29	$4.28	$4.30	$4.32	$4.33	$4.32	$4.31	$4.29	$4.25	$4.22
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	198	199	200	202	203	203	204	204	203	202	202	202	202	203	205	207

Drivers of Demand Data

	May-2007	Jun-2007	Jul-2007	Aug-2007	Sep-2007	Oct-2007	Nov-2007	Dec-2007	Jan-2008	Feb-2008	Mar-2008	Apr-2008	May-2008	Jun-2008	Jul-2008	Aug-2008
Population & Households																
Population	777,003	777,838	778,673	779,713	780,753	781,793	782,833	783,873	784,913	785,952	786,992	788,032	789,072	790,112	791,152	792,050
Households	322,047	321,997	321,947	322,063	322,179	322,295	322,411	322,527	322,643	322,759	322,875	322,991	323,107	323,223	323,339	323,443
Housing Units	357,609	357,721	357,833	358,006	358,178	358,351	358,524	358,696	358,869	359,042	359,214	359,387	359,560	359,732	359,905	360,016
Housing Units	367,985	368,038	368,049	368,154	368,749	368,816	369,300	369,659	369,979	370,118	370,378	370,819	371,231	371,332	371,515	371,641
Households, Income $0-15k	40,770	40,491	40,212	40,300	40,389	40,477	40,565	40,654	40,742	40,830	40,919	41,007	41,095	41,184	41,272	41,361
Households, Income $15-50k	84,736	84,708	84,679	83,952	83,226	82,499	81,773	81,046	80,320	79,593	78,866	78,140	77,413	76,687	75,960	76,137
Households, Income $50-100k	85,920	85,612	85,303	85,143	84,982	84,822	84,662	84,501	84,341	84,181	84,020	83,860	83,700	83,539	83,379	83,885
Households, Income $100k+	110,621	111,187	111,753	112,668	113,582	114,497	115,411	116,326	117,241	118,155	119,070	119,984	120,899	121,813	122,728	122,059
Households, 0 Vehicles	94,643	94,878	95,112	95,218	95,324	95,430	95,536	95,642	95,748	95,854	95,960	96,066	96,172	96,278	96,384	96,541
Median Household Income (2010$)	$70,888	$70,971	$71,209	$71,845	$72,151	$72,499	$72,570	$73,119	$73,254	$73,538	$73,397	$73,443	$73,314	$73,062	$73,158	$73,200
Workers (at home location)																
Workers	409,074	410,648	413,540	410,338	414,027	416,301	417,600	417,603	410,490	412,365	411,469	414,505	414,942	414,177	413,189	407,558
Workers, earning $0-15k	93,156	93,060	93,261	92,273	92,834	93,074	93,093	92,822	90,974	91,122	90,656	91,055	90,881	90,444	89,959	88,660
Workers, earning $15-40k	121,050	121,278	121,895	120,669	121,469	121,849	121,942	121,655	119,300	119,560	119,017	119,609	119,449	118,943	118,374	116,550
Workers, earning $40k+	194,868	196,310	198,384	197,396	199,724	201,378	202,566	203,126	200,216	201,683	201,796	203,841	204,612	204,790	204,856	202,348
Employment (at work location)																
Total Employment	552,761	555,042	559,104	556,321	562,883	567,545	570,891	572,470	564,269	568,403	568,722	574,485	576,658	577,161	577,347	570,883
Retail Employment	43,347	43,460	43,862	43,859	44,097	44,253	46,165	47,243	44,660	43,637	43,409	43,799	43,740	43,843	44,042	44,281
Education and Health Employment	98,704	98,325	97,441	92,660	99,330	100,547	101,096	101,036	98,636	101,037	100,966	101,607	101,700	100,372	99,995	94,651
Leisure Employment	77,752	78,466	79,743	79,492	80,032	79,484	79,223	80,053	75,463	77,142	77,269	79,502	80,721	81,172	80,882	80,567
Other Employment	332,958	334,791	338,058	340,310	339,424	343,261	344,407	344,138	345,510	346,587	347,078	349,577	350,497	351,774	352,428	351,384
Employees, earning $0-15k	115,308	115,532	116,126	115,232	116,274	116,919	117,289	117,296	115,304	115,836	115,590	116,449	116,577	116,368	116,096	114,763
Employees, earning $15-40k	155,033	155,124	155,711	154,664	156,215	157,235	157,888	158,051	155,519	156,390	156,210	157,524	157,851	157,722	157,506	155,429
Employees, earning $40k+	282,420	284,386	287,268	286,425	290,394	293,391	295,714	297,123	293,446	296,178	296,922	300,512	302,230	303,071	303,744	300,692
Average monthly earnings (2010$)	$5,497	$5,435	$5,385	$5,736	$6,061	$6,387	$6,486	$6,628	$6,731	$6,308	$5,854	$5,420	$5,398	$5,367	$5,361	$5,572
Jobs-Housing Balance																
Employees per Housing Unit	1.50	1.51	1.52	1.51	1.53	1.54	1.55	1.55	1.53	1.54	1.54	1.55	1.55	1.55	1.55	1.54
Employees per Worker	1.35	1.35	1.35	1.36	1.36	1.36	1.37	1.37	1.37	1.38	1.38	1.39	1.39	1.39	1.40	1.40
Workers: Live & Work in SF	246,220	247,363	249,300	247,371	249,597	250,971	251,756	251,760	247,474	248,606	248,068	249,901	250,166	249,707	249,114	245,823
Workers: Live elswhere & work in SF	305,030	306,130	308,212	307,354	311,661	314,925	317,465	319,025	315,124	318,103	318,948	322,851	324,741	325,691	326,460	323,248
Workers: Live in SF & work elsewhere	161,496	161,895	162,814	161,525	162,949	163,816	164,298	164,271	161,445	162,154	161,773	162,939	163,082	162,753	162,336	160,022
Costs																
Average Fuel Price (2010$)	$3.75	$3.61	$3.45	$3.22	$3.14	$3.31	$3.63	$3.60	$3.52	$3.44	$3.75	$3.94	$4.07	$4.50	$4.49	$4.18
Average Fleet Efficiency (mpg)	22.83	22.87	22.90	22.97	23.03	23.10	23.17	23.23	23.30	23.37	23.43	23.50	23.57	23.63	23.70	23.68
Average Fuel Cost (2010$ / mi)	$0.16	$0.16	$0.15	$0.14	$0.14	$0.14	$0.16	$0.15	$0.15	$0.15	$0.16	$0.17	$0.17	$0.19	$0.19	$0.18
Average Auto Operating Cost (2010$/mile)	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20	$0.20
Median Daily CBD Parking Cost (2010$)			$28.26	$28.32	$28.24	$28.18	$28.01	$28.03	$27.89	$27.81	$27.57	$27.41	$27.18	$26.91	$27.76	$27.87
Median Monthly CBD Parking Cost (2010$)			$366.39	$367.07	$366.06	$365.28	$363.12	$363.36	$361.57	$360.52	$357.42	$355.27	$352.30	$348.79	$346.96	$348.35
Bay Bridge Toll, Peak (2010$)	$4.19	$4.19	$4.19	$4.20	$4.18	$4.17	$4.15	$4.15	$4.13	$4.12	$4.08	$4.06	$4.03	$3.99	$3.97	$3.98
Bay Bridge Toll, Off-Peak (2010$)	$4.19	$4.19	$4.19	$4.20	$4.18	$4.17	$4.15	$4.15	$4.13	$4.12	$4.08	$4.06	$4.03	$3.99	$3.97	$3.98
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$4.19	$4.19	$4.19	$4.20	$4.18	$4.17	$4.15	$4.15	$4.13	$4.12	$4.08	$4.06	$4.03	$3.99	$3.97	$3.98
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	208	208	208	208	208	209	210	210	211	212	214	215	217	219	220	219

Drivers of Demand Data

	Sep-2008	Oct-2008	Nov-2008	Dec-2008	Jan-2009	Feb-2009	Mar-2009	Apr-2009	May-2009	Jun-2009	Jul-2009	Aug-2009	Sep-2009	Oct-2009	Nov-2009	Dec-2009
Population & Households																
Population	792,947	793,845	794,742	795,640	796,537	797,435	798,332	799,230	800,127	801,025	801,922	802,247	802,573	802,898	803,223	803,548
Households	323,547	323,651	323,755	323,859	323,964	324,068	324,172	324,276	324,380	324,484	324,588	325,540	326,492	327,444	328,396	329,348
Housing Units	360,128	360,239	360,351	360,462	360,574	360,685	360,796	360,908	361,019	361,131	361,242	362,537	363,831	365,126	366,420	367,715
Housing Units	371,697	371,949	372,276	372,780	373,242	373,649	373,721	373,884	374,085	374,103	374,139	374,380	375,213	375,628	376,231	376,391
Households, Income $0-15k	41,451	41,540	41,630	41,719	41,809	41,898	41,987	42,077	42,166	42,256	42,345	42,514	42,684	42,853	43,022	43,191
Households, Income $15-50k	76,314	76,492	76,669	76,846	77,023	77,200	77,377	77,555	77,732	77,909	78,086	78,552	79,019	79,485	79,952	80,418
Households, Income $50-100k	84,391	84,898	85,404	85,910	86,416	86,922	87,428	87,935	88,441	88,947	89,453	88,956	88,459	87,962	87,464	86,967
Households, Income $100k+	121,391	120,722	120,053	119,385	118,716	118,047	117,379	116,710	116,041	115,373	114,704	115,518	116,331	117,145	117,958	118,772
Households, 0 Vehicles	96,698	96,854	97,011	97,168	97,325	97,481	97,638	97,795	97,952	98,108	98,265	98,646	99,027	99,409	99,790	100,171
Median Household Income (2010$)	$73,050	$73,541	$74,718	$75,237	$74,651	$74,022	$73,584	$73,142	$72,674	$71,800	$71,659	$71,581	$71,618	$71,631	$71,662	$71,871
Workers (at home location)																
Workers	409,722	408,804	406,564	405,456	393,555	391,618	388,553	386,290	385,567	381,924	376,944	375,136	376,598	378,708	377,859	376,826
Workers, earning $0-15k	89,056	88,782	88,220	87,904	85,251	84,757	84,020	83,456	83,226	82,365	81,217	80,504	80,493	80,619	80,113	79,571
Workers, earning $15-40k	116,956	116,480	115,627	115,096	111,507	110,747	109,669	108,819	108,402	107,165	105,557	104,894	105,145	105,577	105,183	104,739
Workers, earning $40k+	203,710	203,543	202,717	202,455	196,798	196,114	194,864	194,015	193,939	192,394	190,170	189,739	190,959	192,513	192,563	192,516
Employment (at work location)																
Total Employment	575,336	575,477	573,756	573,631	558,202	556,865	553,917	552,105	552,492	548,690	542,947	541,409	544,587	548,712	548,550	548,115
Retail Employment	43,916	44,313	45,085	45,620	43,523	42,572	41,786	41,136	40,593	40,170	40,162	40,477	40,278	40,462	41,954	43,236
Education and Health Employment	100,617	101,497	102,236	102,294	100,905	102,364	101,801	102,399	102,245	100,246	96,101	95,402	100,369	100,782	101,216	100,892
Leisure Employment	81,325	79,964	79,050	79,436	74,393	74,539	73,860	75,658	78,150	77,396	77,624	77,563	77,506	77,391	75,420	75,391
Other Employment	349,478	349,703	347,385	346,281	339,381	337,390	336,470	332,912	331,504	330,878	329,060	327,967	326,434	330,077	329,960	328,596
Employees, earning $0-15k	115,624	115,618	115,238	115,179	112,048	111,746	111,121	110,724	110,769	109,973	108,789	107,761	107,673	107,767	107,017	106,219
Employees, earning $15-40k	156,323	156,043	155,258	154,905	150,427	149,756	148,652	147,856	147,649	146,323	144,484	143,876	144,521	145,416	145,175	144,863
Employees, earning $40k+	303,389	303,816	303,260	303,547	295,727	295,363	294,144	293,525	294,075	292,394	289,674	289,772	292,393	295,529	296,357	297,033
Average monthly earnings (2010$)	$5,770	$6,021	$6,160	$6,247	$6,242	$5,925	$5,625	$5,326	$5,314	$5,273	$5,286	$5,592	$5,906	$6,218	$6,291	$6,380
Jobs-Housing Balance																
Employees per Housing Unit	1.55	1.55	1.54	1.54	1.50	1.49	1.48	1.48	1.48	1.47	1.45	1.45	1.45	1.46	1.46	1.46
Employees per Worker	1.40	1.41	1.41	1.41	1.42	1.42	1.43	1.43	1.43	1.44	1.44	1.44	1.45	1.45	1.45	1.45
Workers: Live & Work in SF	247,233	246,785	245,538	244,975	237,889	236,822	235,073	233,808	233,475	231,374	228,462	227,386	228,292	229,592	229,097	228,491
Workers: Live elswhere & work in SF	326,218	326,747	326,219	326,598	318,253	317,930	316,685	316,086	316,747	315,004	312,141	311,678	313,930	316,730	317,057	317,223
Workers: Live in SF & work elsewhere	160,769	160,306	159,324	158,786	154,024	153,164	151,863	150,877	150,492	148,968	146,924	146,175	146,700	147,478	147,103	146,657
Costs																
Average Fuel Price (2010$)	$3.92	$3.60	$2.61	$1.97	$2.13	$2.31	$2.28	$2.42	$2.58	$3.02	$2.99	$3.12	$3.24	$3.15	$3.06	$3.00
Average Fleet Efficiency (mpg)	23.67	23.65	23.63	23.62	23.60	23.58	23.57	23.55	23.53	23.52	23.50	23.48	23.47	23.45	23.43	23.42
Average Fuel Cost (2010$ / mi)	$0.17	$0.15	$0.11	$0.08	$0.09	$0.10	$0.10	$0.10	$0.11	$0.13	$0.13	$0.13	$0.14	$0.13	$0.13	$0.13
Average Auto Operating Cost (2010$/mile)	$0.20	$0.20	$0.20	$0.20	$0.25	$0.25	$0.25	$0.25	$0.25	$0.25	$0.27	$0.27	$0.27	$0.27	$0.27	$0.27
Median Daily CBD Parking Cost (2010$)	$27.91	$28.19	$28.74	$29.04	$28.92	$28.77	$28.70	$28.63	$28.55	$28.31	$25.31	$25.26	$25.24	$25.22	$25.20	$25.24
Median Monthly CBD Parking Cost (2010$)	$348.84	$352.40	$359.28	$363.03	$361.46	$359.67	$358.80	$357.90	$356.87	$353.83	$354.40	$353.60	$353.38	$353.04	$352.79	$353.41
Bay Bridge Toll, Peak (2010$)	$3.99	$4.03	$4.11	$4.15	$4.13	$4.11	$4.10	$4.09	$4.08	$4.04	$4.05	$4.04	$4.04	$4.03	$4.03	$4.04
Bay Bridge Toll, Off-Peak (2010$)	$3.99	$4.03	$4.11	$4.15	$4.13	$4.11	$4.10	$4.09	$4.08	$4.04	$4.05	$4.04	$4.04	$4.03	$4.03	$4.04
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Golden Gate Bridge Toll, Peak (2010$)	$4.98	$5.03	$5.13	$5.19	$5.16	$5.14	$5.13	$5.11	$5.10	$5.05	$5.06	$5.05	$5.05	$5.04	$5.04	$5.05
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00
Consumer Price Index	219	217	212	210	211	212	213	213	214	216	215	216	216	216	216	216

Drivers of Demand Data

	Jan-2010	Feb-2010	Mar-2010	Apr-2010	May-2010	Jun-2010	Jul-2010	Aug-2010	Sep-2010	Oct-2010	Nov-2010	Dec-2010	Jan-2011	Feb-2011	Mar-2011	Apr-2011
Population & Households																
Population	803,874	804,199	804,524	804,849	805,175	805,500	805,825	806,693	807,561	808,429	809,296	810,164	811,032	811,900	812,768	813,636
Households	330,300	331,252	332,204	333,156	334,108	335,060	336,012	336,570	337,128	337,686	338,243	338,801	339,359	339,917	340,475	341,033
Housing Units	369,010	370,304	371,599	372,893	374,188	375,482	376,777	376,901	377,024	377,148	377,272	377,395	377,519	377,643	377,766	377,890
Housing Units	376,602	376,683	376,865	376,942	377,055	377,094	377,215	377,252	377,278	377,371	377,489	377,698	377,827	377,876	377,880	377,911
Households, Income $0-15k	43,361	43,530	43,699	43,868	44,038	44,207	44,376	44,356	44,337	44,317	44,297	44,277	44,258	44,238	44,218	44,198
Households, Income $15-50k	80,885	81,351	81,817	82,284	82,750	83,217	83,683	83,977	84,270	84,564	84,857	85,151	85,445	85,738	86,032	86,325
Households, Income $50-100k	86,470	85,973	85,476	84,979	84,481	83,984	83,487	83,690	83,893	84,097	84,300	84,503	84,706	84,909	85,112	85,316
Households, Income $100k+	119,585	120,399	121,212	122,026	122,839	123,653	124,466	124,547	124,628	124,709	124,789	124,870	124,951	125,032	125,113	125,194
Households, 0 Vehicles	100,552	100,933	101,314	101,696	102,077	102,458	102,839	102,966	103,093	103,220	103,347	103,474	103,602	103,729	103,856	103,983
Median Household Income (2010$)	$71,708	$71,772	$71,559	$71,517	$71,542	$71,694	$71,760	$71,507	$71,311	$71,069	$70,885	$70,610	$70,123	$69,627	$68,804	$68,214
Workers (at home location)																
Workers	364,764	366,465	367,351	370,428	371,987	370,416	369,665	371,192	371,720	374,967	373,022	375,352	363,662	366,826	367,436	370,601
Workers, earning $0-15k	76,712	76,756	76,628	76,954	76,961	76,322	75,853	76,189	76,319	77,007	76,630	77,130	74,749	75,421	75,567	76,239
Workers, earning $15-40k	101,235	101,556	101,649	102,348	102,626	102,040	101,681	101,878	101,801	102,468	101,716	102,131	98,738	99,383	99,336	99,978
Workers, earning $40k+	186,817	188,153	189,073	191,126	192,400	192,054	192,131	193,125	193,600	195,491	194,676	196,091	190,175	192,022	192,533	194,384
Employment (at work location)																
Total Employment	531,599	535,110	537,436	542,978	546,306	545,035	544,963	548,146	549,854	555,588	553,628	558,010	541,521	547,125	548,924	554,546
Retail Employment	39,570	38,764	38,727	39,162	39,274	39,302	39,527	39,793	39,989	40,533	41,936	43,051	40,489	40,155	40,070	40,546
Education and Health Employment	98,613	101,430	102,535	102,776	102,585	100,107	96,351	99,942	101,878	102,881	103,030	103,304	100,266	103,411	103,857	103,983
Leisure Employment	72,409	73,080	73,653	78,144	77,769	78,878	78,978	79,554	80,211	81,421	78,375	78,864	74,806	75,704	76,056	79,282
Other Employment	321,007	321,836	322,521	322,896	326,678	326,748	330,107	328,857	327,776	330,753	330,287	332,791	325,960	327,855	328,941	330,735
Employees, earning $0-15k	102,330	102,317	102,074	102,435	102,371	101,446	100,749	101,456	101,890	103,070	102,823	103,754	100,800	101,955	102,402	103,563
Employees, earning $15-40k	140,308	141,044	141,467	142,735	143,418	142,895	142,687	143,251	143,429	144,656	143,880	144,753	140,221	141,416	141,626	142,821
Employees, earning $40k+	288,961	291,748	293,895	297,808	300,517	300,694	301,527	303,439	304,535	307,862	306,925	309,503	300,500	303,754	304,895	308,161
Average monthly earnings (2010$)	$6,436	$6,095	$5,733	$5,385	$5,398	$5,421	$5,437	$5,711	$5,988	$6,261	$6,463	$6,656	$6,828	$6,411	$5,969	$5,554
Jobs-Housing Balance																
Employees per Housing Unit	1.41	1.42	1.43	1.44	1.45	1.45	1.44	1.45	1.46	1.47	1.47	1.48	1.43	1.45	1.45	1.47
Employees per Worker	1.46	1.46	1.46	1.47	1.47	1.47	1.47	1.48	1.48	1.48	1.48	1.49	1.49	1.49	1.49	1.50
Workers: Live & Work in SF	221,197	222,248	222,804	224,690	225,656	224,722	224,286	225,499	226,106	228,367	227,466	229,171	222,308	224,517	225,164	227,378
Workers: Live elswhere & work in SF	308,068	310,506	312,259	315,884	318,225	317,887	318,245	320,102	321,097	324,443	323,296	325,853	316,222	319,493	320,541	323,822
Workers: Live in SF & work elsewhere	141,920	142,539	142,841	143,995	144,558	143,904	143,570	143,864	143,771	144,729	143,683	144,285	139,506	140,434	140,382	141,305
Costs																
Average Fuel Price (2010$)	$3.10	$3.02	$3.12	$3.15	$3.15	$3.18	$3.21	$3.22	$3.10	$3.16	$3.22	$3.32	$3.38	$3.55	$3.93	$4.10
Average Fleet Efficiency (mpg)	23.40	23.38	23.37	23.35	23.33	23.32	23.30	23.29	23.28	23.28	23.27	23.26	23.25	23.24	23.23	23.23
Average Fuel Cost (2010$ / mi)	$0.13	$0.13	$0.13	$0.13	$0.14	$0.14	$0.14	$0.14	$0.13	$0.14	$0.14	$0.14	$0.15	$0.15	$0.17	$0.18
Average Auto Operating Cost (2010$/mile)	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.17	$0.19	$0.19	$0.19	$0.19
Median Daily CBD Parking Cost (2010$)	$25.16	$25.15	$25.05	$25.01	$24.99	$25.01	$25.01	$24.97	$24.96	$24.93	$24.91	$24.87	$24.75	$24.63	$24.39	$24.24
Median Monthly CBD Parking Cost (2010$)	$352.21	$352.12	$350.68	$350.08	$349.80	$350.15	$375.08	$374.56	$374.34	$373.88	$373.72	$373.08	$371.31	$369.49	$365.92	$363.58
Bay Bridge Toll, Peak (2010$)	$4.03	$4.02	$4.01	$4.00	$4.00	$4.00	$6.00	$5.99	$5.99	$5.98	$5.98	$5.97	$5.94	$5.91	$5.85	$5.82
Bay Bridge Toll, Off-Peak (2010$)	$4.03	$4.02	$4.01	$4.00	$4.00	$4.00	$4.00	$4.00	$3.99	$3.99	$3.99	$3.98	$3.96	$3.94	$3.90	$3.88
Bay Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$2.50	$2.50	$2.50	$2.49	$2.49	$2.49	$2.48	$2.46	$2.44	$2.42
Golden Gate Bridge Toll, Peak (2010$)	$5.03	$5.03	$5.01	$5.00	$5.00	$5.00	$5.00	$4.99	$4.99	$4.99	$4.98	$4.97	$4.95	$4.93	$4.88	$4.85
Golden Gate Bridge Toll, Carpools (2010$)	$0.00	$0.00	$0.00	$0.00	$0.00	$0.00	$3.00	$3.00	$2.99	$2.99	$2.99	$2.98	$2.97	$2.96	$2.93	$2.91
Consumer Price Index	217	217	218	218	218	218	218	218	218	219	219	219	220	221	223	225

Drivers of Demand Data

	May-2011	Jun-2011	Jul-2011	Aug-2011	Sep-2011	Oct-2011	Nov-2011	Dec-2011	Jan-2012	Feb-2012	Mar-2012	Apr-2012	May-2012	Jun-2012	Jul-2012	Aug-2012
Population & Households																
Population	814,503	815,371	816,239	817,360	818,481	819,602	820,723	821,844	822,965	824,086	825,207	826,328	827,449	828,570	829,691	830,645
Households	341,590	342,148	342,706	343,051	343,395	343,740	344,085	344,429	344,774	345,119	345,463	345,808	346,153	346,497	346,842	347,493
Housing Units	378,014	378,137	378,261	378,148	378,034	377,921	377,807	377,694	377,580	377,467	377,353	377,240	377,126	377,013	376,899	377,241
Housing Units	377,944	377,949	378,102	378,124	378,122	378,089	378,095	378,105	378,102	378,022	378,093	378,148	378,180	378,447	378,549	378,593
Households, Income $0-15k	44,179	44,159	44,139	44,675	45,211	45,747	46,283	46,819	47,355	47,890	48,426	48,962	49,498	50,034	50,570	50,192
Households, Income $15-50k	86,619	86,912	87,206	86,538	85,870	85,202	84,533	83,865	83,197	82,529	81,861	81,193	80,524	79,856	79,188	79,386
Households, Income $50-100k	85,519	85,722	85,925	85,674	85,424	85,173	84,922	84,671	84,421	84,170	83,919	83,668	83,418	83,167	82,916	82,801
Households, Income $100k+	125,274	125,355	125,436	126,164	126,891	127,619	128,347	129,074	129,802	130,530	131,257	131,985	132,713	133,440	134,168	135,114
Households, 0 Vehicles	104,110	104,237	104,364	104,751	105,139	105,526	105,914	106,301	106,689	107,076	107,463	107,851	108,238	108,626	109,013	108,590
Median Household Income (2010$)	$67,746	$67,669	$67,460	$67,525	$67,672	$68,063	$68,370	$68,791	$68,739	$68,687	$68,416	$68,456	$68,783	$69,131	$69,491	$69,459
Workers (at home location)																
Workers	372,145	371,336	370,290	373,954	375,938	376,801	377,110	377,301	371,205	374,997	376,141	380,191	382,598	383,597	380,861	385,163
Workers, earning $0-15k	76,578	76,432	76,238	76,846	77,107	77,137	77,053	76,945	75,558	76,184	76,271	76,945	77,285	77,339	76,641	77,237
Workers, earning $15-40k	100,181	99,752	99,261	99,990	100,266	100,241	100,069	99,866	98,003	98,752	98,802	99,611	99,986	99,991	99,024	99,871
Workers, earning $40k+	195,386	195,152	194,791	197,119	198,566	199,423	199,987	200,490	197,644	200,060	201,068	203,634	205,327	206,267	205,196	208,055
Employment (at work location)																
Total Employment	557,748	557,421	556,730	563,275	567,304	569,646	571,153	572,482	564,253	571,046	573,820	581,038	585,763	588,339	585,179	591,945
Retail Employment	40,742	40,899	41,471	41,681	41,581	41,723	43,526	44,417	41,658	41,040	41,083	41,579	41,840	42,024	42,482	42,897
Education and Health Employment	103,493	101,566	97,799	101,683	103,309	104,866	104,442	104,324	102,024	105,161	105,595	106,046	106,436	104,027	101,564	104,650
Leisure Employment	80,975	80,933	81,066	81,718	83,521	82,731	81,206	81,187	77,197	79,056	79,868	83,287	84,587	85,267	85,018	86,744
Other Employment	332,538	334,023	336,394	338,193	338,893	340,326	341,979	342,554	343,374	345,789	347,274	350,126	352,900	357,021	356,115	357,654
Employees, earning $0-15k	104,273	104,323	104,303	105,460	106,146	106,516	106,729	106,909	105,305	106,505	106,955	108,233	109,045	109,456	108,801	109,884
Employees, earning $15-40k	143,392	143,055	142,628	143,918	144,560	144,770	144,767	144,719	142,261	143,594	143,912	145,339	146,137	146,396	145,230	146,715
Employees, earning $40k+	310,083	310,043	309,799	313,896	316,598	318,360	319,657	320,854	316,687	320,947	322,953	327,466	330,581	332,486	331,147	335,346
Average monthly earnings (2010$)	$5,558	$5,595	$5,621	$5,787	$5,959	$6,152	$6,399	$6,657	$6,869	$6,442	$5,999	$5,588	$5,593	$5,600	$5,608	$5,852
Jobs-Housing Balance																
Employees per Housing Unit	1.48	1.47	1.47	1.49	1.50	1.51	1.51	1.51	1.49	1.51	1.52	1.54	1.55	1.55	1.55	1.56
Employees per Worker	1.50	1.50	1.50	1.51	1.51	1.51	1.51	1.52	1.52	1.52	1.53	1.53	1.53	1.53	1.54	1.54
Workers: Live & Work in SF	228,600	228,375	228,003	230,317	231,597	232,186	232,435	232,611	228,910	231,306	232,069	234,626	236,170	236,845	235,214	237,686
Workers: Live elswhere & work in SF	325,689	325,496	325,091	329,239	331,921	333,618	334,826	335,930	331,420	335,731	337,682	342,253	345,360	347,202	345,658	349,835
Workers: Live in SF & work elsewhere	141,608	141,016	140,338	141,660	142,346	142,606	142,657	142,663	140,293	141,660	142,027	143,490	144,332	144,642	143,544	145,316
Costs																
Average Fuel Price (2010$)	$4.10	$3.85	$3.73	$3.71	$3.83	$3.76	$3.70	$3.51	$3.59	$3.86	$4.17	$4.05	$4.14	$3.94	$3.64	$3.91
Average Fleet Efficiency (mpg)	23.22	23.21	23.20	23.21	23.22	23.23	23.23	23.24	23.25	23.26	23.27	23.28	23.28	23.29	23.30	23.31
Average Fuel Cost (2010$ / mi)	$0.18	$0.17	$0.16	$0.16	$0.17	$0.16	$0.16	$0.15	$0.15	$0.17	$0.18	$0.17	$0.18	$0.17	$0.16	$0.17
Average Auto Operating Cost (2010$/mile)	$0.19	$0.19	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22
Median Daily CBD Parking Cost (2010$)	$24.13	$24.15	$25.09	$25.03	$24.99	$25.04	$25.06	$25.12	$25.01	$24.90	$24.72	$24.64	$24.67	$24.71	$27.60	
Median Monthly CBD Parking Cost (2010$)	$361.88	$362.26	$361.94	$360.95	$360.40	$361.15	$361.45	$362.34	$360.76	$359.18	$356.47	$355.39	$355.81	$356.33	$356.92	
Bay Bridge Toll, Peak (2010$)	$5.79	$5.80	$5.79	$5.78	$5.77	$5.78	$5.78	$5.80	$5.77	$5.75	$5.70	$5.69	$5.69	$5.70	$5.71	$5.68
Bay Bridge Toll, Off-Peak (2010$)	$3.86	$3.86	$3.86	$3.85	$3.84	$3.85	$3.86	$3.87	$3.85	$3.83	$3.80	$3.79	$3.80	$3.80	$3.81	$3.79
Bay Bridge Toll, Carpools (2010$)	$2.41	$2.42	$2.41	$2.41	$2.40	$2.41	$2.41	$2.42	$2.41	$2.39	$2.38	$2.37	$2.37	$2.38	$2.38	$2.37
Golden Gate Bridge Toll, Peak (2010$)	$4.83	$4.83	$4.83	$4.81	$4.81	$4.82	$4.82	$4.83	$4.81	$4.79	$4.75	$4.74	$4.74	$4.75	$4.76	$4.73
Golden Gate Bridge Toll, Carpools (2010$)	$2.90	$2.90	$2.90	$2.89	$2.88	$2.89	$2.89	$2.90	$2.89	$2.87	$2.85	$2.84	$2.85	$2.85	$2.86	$2.84
Consumer Price Index	226	226	226	227	227	226	226	226	227	228	229	230	230	229	229	230

Drivers of Demand Data

	Sep-2012	Oct-2012	Nov-2012	Dec-2012	Jan-2013	Feb-2013	Mar-2013	Apr-2013	May-2013	Jun-2013	Jul-2013	Aug-2013	Sep-2013	Oct-2013	Nov-2013	Dec-2013
Population & Households																
Population	831,599	832,553	833,507	834,461	835,415	836,368	837,322	838,276	839,230	840,184	841,138	842,082	843,027	843,971	844,915	845,859
Households	348,144	348,794	349,445	350,096	350,747	351,397	352,048	352,699	353,350	354,000	354,651	354,547	354,444	354,340	354,236	354,132
Housing Units	377,583	377,925	378,266	378,608	378,950	379,292	379,634	379,976	380,317	380,659	381,001	381,468	381,936	382,403	382,871	383,338
Housing Units	378,619	378,656	378,809	379,101	379,423	379,471	379,978	380,064	380,142	380,215	380,284	379,969	380,031	380,407	380,953	381,336
Households, Income $0-15k	49,814	49,436	49,058	48,680	48,302	47,924	47,546	47,168	46,790	46,412	46,034	45,704	45,374	45,044	44,714	44,384
Households, Income $15-50k	79,584	79,782	79,980	80,178	80,377	80,575	80,773	80,971	81,169	81,367	81,565	80,930	80,295	79,659	79,024	78,389
Households, Income $50-100k	82,685	82,570	82,454	82,339	82,224	82,108	81,993	81,877	81,762	81,646	81,531	81,618	81,706	81,793	81,881	81,968
Households, Income $100k+	136,060	137,006	137,952	138,898	139,845	140,791	141,737	142,683	143,629	144,575	145,521	146,295	147,069	147,843	148,617	149,391
Households, 0 Vehicles	108,167	107,744	107,321	106,898	106,476	106,053	105,630	105,207	104,784	104,361	103,938	104,495	105,051	105,608	106,164	106,721
Median Household Income (2010$)	$69,502	$69,880	$70,566	$71,111	$71,254	$71,025	$71,189	$71,613	$71,835	$72,011	$72,330	$72,833	$73,337	$74,116	$74,859	$75,457
Workers (at home location)																
Workers	386,250	389,519	390,598	392,330	386,391	390,846	392,302	394,436	396,262	396,266	396,437	397,955	398,118	399,863	403,444	405,144
Workers, earning $0-15k	77,187	77,572	77,520	77,597	76,162	76,779	76,805	76,963	77,060	76,804	76,581	76,790	76,738	76,991	77,596	77,839
Workers, earning $15-40k	99,883	100,457	100,466	100,643	98,857	99,734	99,843	100,124	100,326	100,067	99,853	99,959	99,726	99,889	100,508	100,657
Workers, earning $40k+	209,181	211,490	212,612	214,089	211,372	214,334	215,655	217,350	218,876	219,395	220,003	221,205	221,654	222,983	225,340	226,647
Employment (at work location)																
Total Employment	593,772	598,953	600,766	603,584	594,599	601,606	603,998	607,434	610,396	610,551	610,963	614,087	615,119	618,596	624,920	628,335
Retail Employment	42,931	43,641	45,488	46,045	43,995	43,593	43,449	43,434	43,484	43,753	44,304	44,605	44,510	44,971	46,931	47,567
Education and Health Employment	106,507	107,521	108,315	108,465	126,909	130,202	130,336	129,834	130,018	127,565	125,763	126,207	127,881	128,594	129,137	129,244
Leisure Employment	86,886	87,635	85,954	85,961	81,631	82,894	83,955	86,190	87,456	86,806	85,614	86,963	87,643	87,366	87,320	87,595
Other Employment	357,448	360,156	361,009	363,113	342,064	344,917	346,258	347,976	349,438	352,427	355,282	356,312	355,085	357,665	361,532	363,929
Employees, earning $0-15k	110,048	110,834	110,995	111,343	109,516	110,636	110,907	111,369	111,743	111,605	111,514	112,030	112,164	112,745	113,843	114,412
Employees, earning $15-40k	146,975	148,064	148,320	148,824	146,422	147,959	148,360	149,017	149,557	149,410	149,328	149,928	150,017	150,703	152,082	152,751
Employees, earning $40k+	336,749	340,055	341,451	343,417	338,662	343,011	344,731	347,048	349,095	349,536	350,121	352,129	352,937	355,148	358,995	361,173
Average monthly earnings (2010$)	$6,100	$6,376	$6,515	$6,642	$6,731	$6,345	$5,999	$5,675	$5,714	$5,749	$5,795	$6,043	$6,290	$6,562	$6,812	$7,050
Jobs-Housing Balance																
Employees per Housing Unit	1.57	1.58	1.59	1.59	1.57	1.59	1.59	1.60	1.61	1.61	1.61	1.62	1.62	1.63	1.64	1.65
Employees per Worker	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.54	1.55	1.55	1.55	1.55
Workers: Live & Work in SF	238,173	240,005	240,486	241,369	237,537	240,097	240,813	241,944	242,886	242,712	242,642	243,283	243,096	243,875	245,773	246,521
Workers: Live elswhere & work in SF	351,094	354,337	355,588	357,434	352,288	356,614	358,206	360,417	362,348	362,611	363,027	365,448	366,623	369,256	373,592	376,193
Workers: Live in SF & work elsewhere	145,877	147,262	147,820	148,624	146,521	148,357	149,055	150,012	150,852	150,997	151,206	152,051	152,377	153,309	154,948	155,865
Costs																
Average Fuel Price (2010$)	$3.99	$4.22	$3.72	$3.46	$3.48	$3.83	$3.92	$3.80	$3.82	$3.80	$3.78	$3.67	$3.73	$3.66	$3.45	$3.42
Average Fleet Efficiency (mpg)	23.32	23.33	23.33	23.34	23.35	23.36	23.37	23.38	23.38	23.39	23.40	23.41	23.42	23.43	23.43	23.44
Average Fuel Cost (2010$ / mi)	$0.17	$0.18	$0.16	$0.15	$0.15	$0.16	$0.17	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.16	$0.15	$0.15
Average Auto Operating Cost (2010$/mile)	$0.22	$0.22	$0.22	$0.22	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23	$0.23
Median Daily CBD Parking Cost (2010$)																
Median Monthly CBD Parking Cost (2010$)																
Bay Bridge Toll, Peak (2010$)	$5.65	$5.66	$5.68	$5.70	$5.68	$5.64	$5.62	$5.63	$5.62	$5.60	$5.60	$5.59	$5.59	$5.60	$5.61	$5.61
Bay Bridge Toll, Off-Peak (2010$)	$3.77	$3.77	$3.79	$3.80	$3.79	$3.76	$3.75	$3.75	$3.74	$3.74	$3.73	$3.73	$3.73	$3.73	$3.74	$3.74
Bay Bridge Toll, Carpools (2010$)	$2.36	$2.36	$2.37	$2.37	$2.37	$2.35	$2.34	$2.34	$2.34	$2.33	$2.33	$2.33	$2.33	$2.33	$2.34	$2.34
Golden Gate Bridge Toll, Peak (2010$)	$4.71	$4.71	$4.74	$4.75	$4.73	$4.70	$4.68	$4.69	$4.68	$4.67	$4.67	$4.66	$4.66	$4.67	$4.68	$4.68
Golden Gate Bridge Toll, Carpools (2010$)	$2.83	$2.83	$2.84	$2.85	$2.84	$2.82	$2.81	$2.81	$2.81	$2.80	$2.80	$2.80	$2.79	$2.80	$2.81	$2.81
Consumer Price Index	231	231	230	230	230	232	233	233	233	234	234	234	234	234	233	233

Drivers of Demand Data

	Jan-2014	Feb-2014	Mar-2014	Apr-2014	May-2014	Jun-2014	Jul-2014	Aug-2014	Sep-2014	Oct-2014	Nov-2014	Dec-2014	Jan-2015	Feb-2015	Mar-2015
Population & Households															
Population	846,804	847,748	848,692	849,636	850,581	851,525	852,469	853,413	854,358	855,302	856,246	857,190			
Households	354,029	353,925	353,821	353,717	353,614	353,510	353,406	353,302	353,199	353,095	352,991	352,887			
Housing Units	383,806	384,273	384,740	385,208	385,675	386,143	386,610	387,077	387,545	388,012	388,480	388,947			
Housing Units	381,382	381,375	381,428	381,520	381,562	381,576	382,187	382,634	382,754	382,837	383,541	384,661			
Households, Income $0-15k	44,055	43,725	43,395	43,065	42,735	42,405	42,075	41,745	41,415	41,085	40,755	40,425			
Households, Income $15-50k	77,754	77,118	76,483	75,848	75,213	74,577	73,942	73,307	72,672	72,036	71,401	70,766			
Households, Income $50-100k	82,056	82,143	82,230	82,318	82,405	82,493	82,580	82,667	82,755	82,842	82,930	83,017			
Households, Income $100k+	150,165	150,939	151,713	152,487	153,261	154,035	154,809	155,583	156,357	157,131	157,905	158,679			
Households, 0 Vehicles	107,278	107,834	108,391	108,947	109,504	110,060	110,617	111,174	111,730	112,287	112,843	113,400			
Median Household Income (2010$)	$75,767	$76,075	$76,171	$76,502	$76,815	$77,251	$77,859	$78,569	$79,089	$79,869	$80,886	$81,934			
Workers (at home location)															
Workers	398,532	402,222	404,001	406,814	408,727	408,638	409,011	412,184	413,705	415,718	416,628	418,913			
Workers, earning $0-15k	76,487	77,113	77,372	77,828	78,111	78,013	78,002	78,526	78,734	79,035	79,127	79,479			
Workers, earning $15-40k	98,745	99,390	99,560	99,983	100,183	99,892	99,716	100,222	100,324	100,544	100,497	100,782			
Workers, earning $40k+	223,299	225,719	227,070	229,003	230,432	230,733	231,293	233,437	234,648	236,139	237,004	238,652			
Employment (at work location)															
Total Employment	618,846	625,344	628,879	634,026	637,776	638,402	639,747	645,473	648,617	652,535	654,722	659,074	650,159	655,385	660,068
Retail Employment	44,859	44,379	44,302	44,759	45,044	45,124	45,717	46,146	45,794	46,074	47,488	48,368	46,577	45,787	46,121
Education and Health Employment	126,761	129,640	129,934	130,446	130,541	127,891	124,814	127,112	130,387	130,754	131,725	131,686	129,913	132,686	133,041
Leisure Employment	84,556	85,963	86,912	89,718	90,192	91,046	90,513	91,449	92,149	91,902	89,924	90,458	87,644	89,347	90,113
Other Employment	362,670	365,362	367,731	369,103	371,999	374,341	378,703	380,766	380,287	383,805	385,585	388,562	386,025	387,565	390,793
Employees, earning $0-15k	112,631	113,761	114,352	115,235	115,864	115,926	116,118	117,105	117,624	118,283	118,628	119,365			
Employees, earning $15-40k	150,286	151,705	152,404	153,493	154,243	154,237	154,406	155,631	156,233	157,021	157,392	158,284			
Employees, earning $40k+	355,929	359,878	362,123	365,298	367,669	368,239	369,223	372,736	374,760	377,231	378,702	381,426			
Average monthly earnings (2010$)	$7,260	$6,794	$6,313	$5,857	$5,952	$6,055	$6,172	$6,383	$6,579	$6,796	$7,104	$7,417	$7,726	$7,692	$7,647
Jobs-Housing Balance															
Employees per Housing Unit	1.62	1.64	1.65	1.66	1.67	1.67	1.67	1.69	1.69	1.70	1.71	1.71			
Employees per Worker	1.55	1.55	1.56	1.56	1.56	1.56	1.56	1.57	1.57	1.57	1.57	1.57			
Workers: Live & Work in SF	242,218	244,178	244,977	246,401	247,278	246,944	246,890	248,526	249,164	250,097	250,366	251,461			
Workers: Live elswhere & work in SF	371,060	375,504	378,175	381,818	384,623	385,544	386,896	390,900	393,344	396,259	398,123	401,306			
Workers: Live in SF & work elsewhere	153,581	155,263	156,210	157,557	158,558	158,783	159,186	160,679	161,530	162,574	163,187	164,339			
Costs															
Average Fuel Price (2010$)	$3.44	$3.47	$3.68	$3.87	$3.90	$3.84	$3.79	$3.68	$3.56	$3.34	$3.05	$2.77	$2.50	$2.58	$3.14
Average Fleet Efficiency (mpg)	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44	23.44
Average Fuel Cost (2010$ / mi)	$0.15	$0.15	$0.16	$0.16	$0.17	$0.16	$0.16	$0.16	$0.15	$0.14	$0.13	$0.12	$0.11	$0.11	$0.13
Average Auto Operating Cost (2010$/mile)	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.22	$0.21	$0.21	$0.21
Median Daily CBD Parking Cost (2010$)															
Median Monthly CBD Parking Cost (2010$)															
Bay Bridge Toll, Peak (2010$)	$5.59	$5.57	$5.54	$5.52	$5.50	$5.49	$5.49	$5.50	$5.50	$5.51	$5.54	$5.57	$5.60	$5.57	$5.54
Bay Bridge Toll, Off-Peak (2010$)	$3.73	$3.72	$3.69	$3.68	$3.67	$3.66	$3.66	$3.67	$3.66	$3.67	$3.69	$3.71	$3.73	$3.72	$3.69
Bay Bridge Toll, Carpools (2010$)	$2.33	$2.32	$2.31	$2.30	$2.29	$2.29	$2.29	$2.29	$2.29	$2.30	$2.31	$2.32	$2.33	$2.32	$2.31
Golden Gate Bridge Toll, Peak (2010$)	$4.66	$4.64	$4.61	$5.52	$5.50	$5.49	$5.49	$5.50	$5.50	$5.51	$5.54	$5.57	$5.60	$5.57	$5.54
Golden Gate Bridge Toll, Carpools (2010$)	$2.80	$2.79	$2.77	$3.68	$3.67	$3.66	$3.66	$3.67	$3.66	$3.67	$3.69	$3.71	$3.73	$3.72	$3.69
Consumer Price Index	234	235	236	237	238	238	238	238	238	237	236	235			

Commute Mode Share Data

	Source	Temporal Res	Geog Res	Trend	Jan-2001	Feb-2001	Mar-2001	Apr-2001	May-2001	Jun-2001	Jul-2001	Aug-2001	Sep-2001
Commute Mode Shares													
Drive-Alone	ACS	Annual	County		40.4%	40.4%	40.4%	40.4%	40.4%	40.3%	40.3%	40.3%	40.3%
Carpool	ACS	Annual	County		10.5%	10.4%	10.4%	10.3%	10.3%	10.3%	10.2%	10.2%	10.2%
Transit	ACS	Annual	County		31.0%	31.0%	31.1%	31.1%	31.1%	31.2%	31.2%	31.2%	31.2%
Walk	ACS	Annual	County		9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%
Taxi, bike, other	ACS	Annual	County		3.9%	3.9%	3.9%	3.9%	3.9%	3.8%	3.8%	3.8%	3.8%
Work at home	ACS	Annual	County		4.8%	4.9%	4.9%	4.9%	4.9%	5.0%	5.0%	5.0%	5.0%
Commute Mode Shares by Segment													
Workers earning $0-50k: Drive-Alone	ACS	Annual	County		39.2%	39.1%	39.0%	39.0%	38.9%	38.8%	38.8%	38.7%	38.6%
Workers earning $0-50k: Carpool	ACS	Annual	County		12.1%	12.0%	12.0%	11.9%	11.9%	11.8%	11.8%	11.7%	11.6%
Workers earning $0-50k: Transit	ACS	Annual	County		31.6%	31.6%	31.7%	31.8%	31.9%	31.9%	32.0%	32.1%	32.2%
Workers earning $0-50k: Taxi, walk, bike, other	ACS	Annual	County		13.0%	13.0%	13.0%	13.0%	13.1%	13.1%	13.1%	13.1%	13.1%
Workers earning $0-50k: Work at home	ACS	Annual	County		4.2%	4.2%	4.2%	4.3%	4.3%	4.3%	4.4%	4.4%	4.4%
Workers earning $50k+: Drive-Alone	ACS	Annual	County		49.0%	49.0%	48.9%	48.9%	48.8%	48.8%	48.8%	48.7%	48.7%
Workers earning $50k+: Carpool	ACS	Annual	County		9.8%	9.7%	9.7%	9.7%	9.6%	9.6%	9.6%	9.5%	9.5%
Workers earning $50k+: Transit	ACS	Annual	County		24.5%	24.6%	24.6%	24.7%	24.7%	24.8%	24.8%	24.9%	24.9%
Workers earning $50k+: Taxi, walk, bike, other	ACS	Annual	County		11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%
Workers earning $50k+: Work at home	ACS	Annual	County		5.4%	5.5%	5.5%	5.5%	5.5%	5.5%	5.5%	5.5%	5.6%
Workers with 0 vehicles: Drive-Alone	ACS	Annual	County		11.8%	11.7%	11.7%	11.7%	11.6%	11.6%	11.5%	11.5%	11.4%
Workers with 0 vehicles: Carpool	ACS	Annual	County		4.2%	4.2%	4.1%	4.1%	4.0%	4.0%	4.0%	3.9%	3.9%
Workers with 0 vehicles: Transit	ACS	Annual	County		52.5%	52.6%	52.6%	52.7%	52.8%	52.8%	52.9%	53.0%	53.0%
Workers with 0 vehicles: Taxi, walk, bike, other	ACS	Annual	County		27.5%	27.5%	27.4%	27.4%	27.4%	27.4%	27.4%	27.3%	27.3%
Workers with 0 vehicles: Work at home	ACS	Annual	County		4.0%	4.1%	4.1%	4.1%	4.2%	4.2%	4.2%	4.3%	4.3%
Workers with 1+ vehicles: Drive-Alone	ACS	Annual	County		47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%
Workers with 1+ vehicles: Carpool	ACS	Annual	County		11.9%	11.9%	11.8%	11.8%	11.8%	11.7%	11.7%	11.7%	11.6%
Workers with 1+ vehicles: Transit	ACS	Annual	County		26.2%	26.2%	26.2%	26.2%	26.2%	26.3%	26.3%	26.3%	26.3%
Workers with 1+ vehicles: Taxi, walk, bike, other	ACS	Annual	County		9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%
Workers with 1+ vehicles: Work at home	ACS	Annual	County		5.0%	5.0%	5.0%	5.1%	5.1%	5.1%	5.1%	5.2%	5.2%

Commute Mode Share Data

	Oct-2001	Nov-2001	Dec-2001	Jan-2002	Feb-2002	Mar-2002	Apr-2002	May-2002	Jun-2002	Jul-2002	Aug-2002	Sep-2002	Oct-2002	Nov-2002	Dec-2002
Commute Mode Shares															
Drive-Alone	40.3%	40.3%	40.3%	40.3%	40.3%	40.2%	40.2%	40.2%	40.2%	40.2%	40.2%	40.2%	40.2%	40.1%	40.1%
Carpool	10.1%	10.1%	10.0%	10.0%	10.0%	9.9%	9.9%	9.9%	9.8%	9.8%	9.7%	9.7%	9.7%	9.6%	9.6%
Transit	31.3%	31.3%	31.3%	31.4%	31.4%	31.4%	31.5%	31.5%	31.5%	31.5%	31.6%	31.6%	31.6%	31.7%	31.7%
Walk	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.4%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%
Taxi, bike, other	3.8%	3.8%	3.8%	3.8%	3.8%	3.8%	3.8%	3.7%	3.7%	3.7%	3.7%	3.7%	3.7%	3.7%	3.7%
Work at home	5.1%	5.1%	5.1%	5.2%	5.2%	5.2%	5.2%	5.3%	5.3%	5.3%	5.3%	5.4%	5.4%	5.4%	5.4%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	38.5%	38.5%	38.4%	38.3%	38.3%	38.2%	38.1%	38.0%	37.9%	37.9%	37.8%	37.7%	37.6%	37.5%	37.5%
Workers earning $0-50k: Carpool	11.6%	11.5%	11.5%	11.4%	11.4%	11.3%	11.2%	11.2%	11.1%	11.0%	11.0%	10.9%	10.9%	10.8%	10.7%
Workers earning $0-50k: Transit	32.2%	32.3%	32.4%	32.5%	32.6%	32.6%	32.7%	32.8%	32.9%	33.0%	33.1%	33.2%	33.2%	33.3%	33.4%
Workers earning $0-50k: Taxi, walk, bike, other	13.1%	13.2%	13.2%	13.2%	13.2%	13.2%	13.3%	13.3%	13.3%	13.3%	13.3%	13.3%	13.4%	13.4%	13.4%
Workers earning $0-50k: Work at home	4.5%	4.5%	4.5%	4.6%	4.6%	4.7%	4.7%	4.7%	4.8%	4.8%	4.8%	4.9%	4.9%	5.0%	5.0%
Workers earning $50k+: Drive-Alone	48.7%	48.6%	48.6%	48.5%	48.5%	48.5%	48.4%	48.4%	48.4%	48.3%	48.3%	48.3%	48.2%	48.2%	48.2%
Workers earning $50k+: Carpool	9.5%	9.4%	9.4%	9.4%	9.3%	9.3%	9.3%	9.3%	9.2%	9.2%	9.2%	9.1%	9.1%	9.1%	9.0%
Workers earning $50k+: Transit	25.0%	25.0%	25.1%	25.2%	25.2%	25.3%	25.3%	25.4%	25.4%	25.5%	25.5%	25.6%	25.6%	25.6%	25.7%
Workers earning $50k+: Taxi, walk, bike, other	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%	11.3%
Workers earning $50k+: Work at home	5.6%	5.6%	5.6%	5.6%	5.6%	5.6%	5.7%	5.7%	5.7%	5.7%	5.7%	5.7%	5.7%	5.7%	5.8%
Workers with 0 vehicles: Drive-Alone	11.4%	11.4%	11.3%	11.3%	11.2%	11.2%	11.1%	11.1%	11.1%	11.0%	11.0%	10.9%	10.9%	10.9%	10.8%
Workers with 0 vehicles: Carpool	3.9%	3.8%	3.8%	3.7%	3.7%	3.7%	3.6%	3.6%	3.6%	3.5%	3.5%	3.4%	3.4%	3.4%	3.3%
Workers with 0 vehicles: Transit	53.1%	53.2%	53.2%	53.3%	53.4%	53.4%	53.5%	53.6%	53.6%	53.7%	53.8%	53.8%	53.9%	53.9%	54.0%
Workers with 0 vehicles: Taxi, walk, bike, other	27.3%	27.3%	27.3%	27.2%	27.2%	27.2%	27.2%	27.2%	27.1%	27.1%	27.1%	27.1%	27.1%	27.1%	27.0%
Workers with 0 vehicles: Work at home	4.3%	4.4%	4.4%	4.4%	4.5%	4.5%	4.5%	4.6%	4.6%	4.6%	4.7%	4.7%	4.7%	4.8%	4.8%
Workers with 1+ vehicles: Drive-Alone	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%	47.1%
Workers with 1+ vehicles: Carpool	11.6%	11.6%	11.5%	11.5%	11.4%	11.4%	11.4%	11.3%	11.3%	11.3%	11.2%	11.2%	11.1%	11.1%	11.1%
Workers with 1+ vehicles: Transit	26.3%	26.3%	26.3%	26.3%	26.3%	26.4%	26.4%	26.4%	26.4%	26.4%	26.4%	26.4%	26.4%	26.5%	26.5%
Workers with 1+ vehicles: Taxi, walk, bike, other	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.7%	9.7%	9.7%
Workers with 1+ vehicles: Work at home	5.2%	5.2%	5.3%	5.3%	5.3%	5.3%	5.4%	5.4%	5.4%	5.5%	5.5%	5.5%	5.5%	5.6%	5.6%

Commute Mode Share Data

	Jan-2003	Feb-2003	Mar-2003	Apr-2003	May-2003	Jun-2003	Jul-2003	Aug-2003	Sep-2003	Oct-2003	Nov-2003	Dec-2003	Jan-2004	Feb-2004	Mar-2004
Commute Mode Shares															
Drive-Alone	40.1%	40.1%	40.1%	40.1%	40.1%	40.1%	40.1%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%	40.0%
Carpool	9.5%	9.5%	9.5%	9.4%	9.4%	9.3%	9.3%	9.3%	9.2%	9.2%	9.1%	9.1%	9.1%	9.0%	9.0%
Transit	31.7%	31.8%	31.8%	31.8%	31.8%	31.9%	31.9%	31.9%	32.0%	32.0%	32.0%	32.1%	32.1%	32.1%	32.2%
Walk	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%
Taxi, bike, other	3.7%	3.7%	3.7%	3.7%	3.6%	3.6%	3.6%	3.6%	3.6%	3.6%	3.6%	3.6%	3.6%	3.6%	3.5%
Work at home	5.5%	5.5%	5.5%	5.5%	5.6%	5.6%	5.6%	5.7%	5.7%	5.7%	5.7%	5.8%	5.8%	5.8%	5.8%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	37.4%	37.3%	37.2%	37.1%	37.0%	37.0%	36.9%	36.8%	36.7%	36.6%	36.5%	36.4%	36.3%	36.2%	36.1%
Workers earning $0-50k: Carpool	10.7%	10.6%	10.5%	10.5%	10.4%	10.3%	10.3%	10.2%	10.1%	10.0%	10.0%	9.9%	9.8%	9.7%	9.7%
Workers earning $0-50k: Transit	33.5%	33.6%	33.7%	33.8%	33.9%	34.0%	34.1%	34.2%	34.3%	34.4%	34.5%	34.6%	34.7%	34.8%	34.9%
Workers earning $0-50k: Taxi, walk, bike, other	13.4%	13.4%	13.5%	13.5%	13.5%	13.5%	13.5%	13.6%	13.6%	13.6%	13.6%	13.6%	13.7%	13.7%	13.7%
Workers earning $0-50k: Work at home	5.0%	5.1%	5.1%	5.2%	5.2%	5.2%	5.3%	5.3%	5.4%	5.4%	5.4%	5.5%	5.5%	5.6%	5.6%
Workers earning $50k+: Drive-Alone	48.1%	48.1%	48.1%	48.0%	48.0%	48.0%	47.9%	47.9%	47.9%	47.8%	47.8%	47.8%	47.7%	47.7%	47.7%
Workers earning $50k+: Carpool	9.0%	9.0%	9.0%	8.9%	8.9%	8.9%	8.8%	8.8%	8.8%	8.8%	8.7%	8.7%	8.7%	8.7%	8.6%
Workers earning $50k+: Transit	25.7%	25.8%	25.8%	25.9%	25.9%	26.0%	26.0%	26.1%	26.1%	26.2%	26.2%	26.3%	26.3%	26.3%	26.4%
Workers earning $50k+: Taxi, walk, bike, other	11.3%	11.3%	11.3%	11.3%	11.3%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%
Workers earning $50k+: Work at home	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%	5.9%
Workers with 0 vehicles: Drive-Alone	10.8%	10.7%	10.7%	10.6%	10.6%	10.6%	10.5%	10.5%	10.4%	10.4%	10.4%	10.3%	10.3%	10.2%	10.2%
Workers with 0 vehicles: Carpool	3.3%	3.3%	3.2%	3.2%	3.2%	3.1%	3.1%	3.0%	3.0%	3.0%	2.9%	2.9%	2.9%	2.8%	2.8%
Workers with 0 vehicles: Transit	54.1%	54.1%	54.2%	54.3%	54.3%	54.4%	54.5%	54.5%	54.6%	54.7%	54.7%	54.8%	54.9%	54.9%	55.0%
Workers with 0 vehicles: Taxi, walk, bike, other	27.0%	27.0%	27.0%	27.0%	26.9%	26.9%	26.9%	26.9%	26.9%	26.8%	26.8%	26.8%	26.8%	26.8%	26.7%
Workers with 0 vehicles: Work at home	4.8%	4.9%	4.9%	4.9%	5.0%	5.0%	5.0%	5.1%	5.1%	5.1%	5.2%	5.2%	5.2%	5.3%	5.3%
Workers with 1+ vehicles: Drive-Alone	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%
Workers with 1+ vehicles: Carpool	11.0%	11.0%	10.9%	10.9%	10.9%	10.8%	10.8%	10.8%	10.7%	10.7%	10.6%	10.6%	10.6%	10.5%	10.5%
Workers with 1+ vehicles: Transit	26.5%	26.5%	26.5%	26.5%	26.5%	26.5%	26.5%	26.6%	26.6%	26.6%	26.6%	26.6%	26.6%	26.6%	26.6%
Workers with 1+ vehicles: Taxi, walk, bike, other	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%
Workers with 1+ vehicles: Work at home	5.6%	5.6%	5.7%	5.7%	5.7%	5.7%	5.8%	5.8%	5.8%	5.9%	5.9%	5.9%	5.9%	6.0%	6.0%

Commute Mode Share Data

	Apr-2004	May-2004	Jun-2004	Jul-2004	Aug-2004	Sep-2004	Oct-2004	Nov-2004	Dec-2004	Jan-2005	Feb-2005	Mar-2005	Apr-2005	May-2005	Jun-2005
Commute Mode Shares															
Drive-Alone	39.9%	39.9%	39.9%	39.9%	39.9%	39.9%	39.9%	39.9%	39.8%	39.8%	39.8%	39.8%	39.8%	39.8%	39.8%
Carpool	8.9%	8.9%	8.8%	8.8%	8.8%	8.7%	8.7%	8.6%	8.6%	8.6%	8.5%	8.5%	8.4%	8.4%	8.3%
Transit	32.2%	32.2%	32.3%	32.3%	32.3%	32.4%	32.4%	32.4%	32.4%	32.5%	32.5%	32.5%	32.6%	32.6%	32.6%
Walk	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.5%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%
Taxi, bike, other	3.5%	3.5%	3.5%	3.5%	3.5%	3.5%	3.5%	3.5%	3.5%	3.5%	3.4%	3.4%	3.4%	3.4%	3.4%
Work at home	5.9%	5.9%	5.9%	6.0%	6.0%	6.0%	6.0%	6.1%	6.1%	6.1%	6.2%	6.2%	6.2%	6.2%	6.3%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	36.0%	35.9%	35.8%	35.7%	35.6%	35.5%	35.4%	35.3%	35.2%	35.1%	35.0%	34.9%	34.8%	34.7%	34.6%
Workers earning $0-50k: Carpool	9.6%	9.5%	9.4%	9.4%	9.3%	9.2%	9.1%	9.0%	8.9%	8.9%	8.8%	8.7%	8.6%	8.5%	8.4%
Workers earning $0-50k: Transit	35.0%	35.1%	35.2%	35.3%	35.4%	35.5%	35.6%	35.7%	35.9%	36.0%	36.1%	36.2%	36.3%	36.4%	36.6%
Workers earning $0-50k: Taxi, walk, bike, other	13.7%	13.8%	13.8%	13.8%	13.8%	13.9%	13.9%	13.9%	13.9%	14.0%	14.0%	14.0%	14.0%	14.1%	14.1%
Workers earning $0-50k: Work at home	5.7%	5.7%	5.8%	5.8%	5.9%	5.9%	6.0%	6.0%	6.1%	6.1%	6.2%	6.2%	6.3%	6.3%	6.4%
Workers earning $50k+: Drive-Alone	47.6%	47.6%	47.6%	47.5%	47.5%	47.5%	47.5%	47.4%	47.4%	47.4%	47.3%	47.3%	47.3%	47.2%	47.2%
Workers earning $50k+: Carpool	8.6%	8.6%	8.6%	8.5%	8.5%	8.5%	8.5%	8.4%	8.4%	8.4%	8.4%	8.3%	8.3%	8.3%	8.3%
Workers earning $50k+: Transit	26.4%	26.5%	26.5%	26.6%	26.6%	26.6%	26.7%	26.7%	26.8%	26.8%	26.9%	26.9%	26.9%	27.0%	27.0%
Workers earning $50k+: Taxi, walk, bike, other	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%	11.4%
Workers earning $50k+: Work at home	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.0%	6.1%	6.1%	6.1%	6.1%	6.1%	6.1%
Workers with 0 vehicles: Drive-Alone	10.1%	10.1%	10.1%	10.0%	10.0%	9.9%	9.9%	9.9%	9.8%	9.8%	9.7%	9.7%	9.7%	9.6%	9.6%
Workers with 0 vehicles: Carpool	2.8%	2.7%	2.7%	2.6%	2.6%	2.6%	2.5%	2.5%	2.5%	2.4%	2.4%	2.4%	2.3%	2.3%	2.3%
Workers with 0 vehicles: Transit	55.0%	55.1%	55.2%	55.2%	55.3%	55.4%	55.4%	55.5%	55.6%	55.6%	55.7%	55.7%	55.8%	55.9%	55.9%
Workers with 0 vehicles: Taxi, walk, bike, other	26.7%	26.7%	26.7%	26.7%	26.7%	26.6%	26.6%	26.6%	26.6%	26.6%	26.5%	26.5%	26.5%	26.5%	26.5%
Workers with 0 vehicles: Work at home	5.3%	5.4%	5.4%	5.4%	5.4%	5.5%	5.5%	5.5%	5.6%	5.6%	5.6%	5.7%	5.7%	5.7%	5.8%
Workers with 1+ vehicles: Drive-Alone	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.2%	47.3%
Workers with 1+ vehicles: Carpool	10.4%	10.4%	10.3%	10.3%	10.3%	10.2%	10.2%	10.1%	10.1%	10.1%	10.0%	10.0%	9.9%	9.9%	9.8%
Workers with 1+ vehicles: Transit	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%	26.7%	26.8%	26.8%	26.8%	26.8%	26.8%	26.8%	26.8%	26.8%
Workers with 1+ vehicles: Taxi, walk, bike, other	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.7%	9.6%	9.6%	9.6%
Workers with 1+ vehicles: Work at home	6.0%	6.0%	6.1%	6.1%	6.1%	6.2%	6.2%	6.2%	6.2%	6.3%	6.3%	6.3%	6.4%	6.4%	6.4%

Commute Mode Share Data

	Jul-2005	Aug-2005	Sep-2005	Oct-2005	Nov-2005	Dec-2005	Jan-2006	Feb-2006	Mar-2006	Apr-2006	May-2006	Jun-2006	Jul-2006	Aug-2006	Sep-2006
Commute Mode Shares															
Drive-Alone	39.7%	39.8%	39.9%	39.9%	40.0%	40.1%	40.1%	40.2%	40.2%	40.3%	40.4%	40.4%	40.5%	40.3%	40.2%
Carpool	8.3%	8.2%	8.2%	8.1%	8.1%	8.1%	8.0%	8.0%	7.9%	7.9%	7.8%	7.8%	7.7%	7.7%	7.6%
Transit	32.7%	32.5%	32.3%	32.1%	31.9%	31.7%	31.5%	31.3%	31.1%	30.9%	30.7%	30.5%	30.3%	30.5%	30.7%
Walk	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%
Taxi, bike, other	3.4%	3.5%	3.6%	3.6%	3.7%	3.8%	3.9%	4.0%	4.0%	4.1%	4.2%	4.3%	4.3%	4.4%	4.4%
Work at home	6.3%	6.4%	6.5%	6.6%	6.7%	6.8%	6.9%	7.0%	7.1%	7.2%	7.3%	7.4%	7.5%	7.5%	7.4%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	34.4%	34.5%	34.6%	34.7%	34.7%	34.8%	34.9%	34.9%	35.0%	35.1%	35.1%	35.2%	35.3%	35.1%	34.9%
Workers earning $0-50k: Carpool	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.2%	8.2%	8.2%	8.2%	8.2%	8.1%	8.0%
Workers earning $0-50k: Transit	36.7%	36.5%	36.2%	36.0%	35.8%	35.6%	35.3%	35.1%	34.9%	34.7%	34.4%	34.2%	34.0%	34.2%	34.5%
Workers earning $0-50k: Taxi, walk, bike, other	14.1%	14.2%	14.4%	14.5%	14.6%	14.7%	14.9%	15.0%	15.1%	15.3%	15.4%	15.5%	15.7%	15.6%	15.6%
Workers earning $0-50k: Work at home	6.4%	6.5%	6.5%	6.5%	6.6%	6.6%	6.7%	6.7%	6.7%	6.8%	6.8%	6.9%	6.9%	7.0%	7.0%
Workers earning $50k+: Drive-Alone	47.2%	47.2%	47.1%	47.1%	47.1%	47.1%	47.0%	47.0%	47.0%	47.0%	46.9%	46.9%	46.9%	46.7%	46.6%
Workers earning $50k+: Carpool	8.2%	8.1%	8.0%	7.9%	7.8%	7.8%	7.7%	7.6%	7.5%	7.4%	7.3%	7.2%	7.1%	7.2%	7.2%
Workers earning $50k+: Transit	27.1%	26.9%	26.8%	26.7%	26.6%	26.5%	26.4%	26.3%	26.1%	26.0%	25.9%	25.8%	25.7%	26.0%	26.2%
Workers earning $50k+: Taxi, walk, bike, other	11.4%	11.4%	11.5%	11.5%	11.6%	11.6%	11.7%	11.7%	11.7%	11.8%	11.8%	11.9%	11.9%	12.0%	12.2%
Workers earning $50k+: Work at home	6.1%	6.3%	6.5%	6.7%	6.9%	7.1%	7.3%	7.5%	7.6%	7.8%	8.0%	8.2%	8.3%	8.1%	7.9%
Workers with 0 vehicles: Drive-Alone	9.5%	9.4%	9.2%	9.0%	8.8%	8.6%	8.4%	8.2%	8.0%	7.8%	7.6%	7.4%	7.2%	7.2%	7.3%
Workers with 0 vehicles: Carpool	2.2%	2.3%	2.4%	2.5%	2.6%	2.7%	2.8%	2.9%	3.0%	3.1%	3.2%	3.3%	3.4%	3.3%	3.1%
Workers with 0 vehicles: Transit	56.0%	55.7%	55.4%	55.1%	54.7%	54.4%	54.1%	53.8%	53.4%	53.1%	52.8%	52.4%	52.1%	52.1%	52.2%
Workers with 0 vehicles: Taxi, walk, bike, other	26.5%	26.8%	27.2%	27.5%	27.9%	28.3%	28.7%	29.1%	29.4%	29.8%	30.2%	30.6%	31.0%	31.3%	31.5%
Workers with 0 vehicles: Work at home	5.8%	5.8%	5.9%	5.9%	5.9%	6.0%	6.0%	6.1%	6.1%	6.1%	6.2%	6.2%	6.3%	6.1%	5.9%
Workers with 1+ vehicles: Drive-Alone	47.3%	47.3%	47.4%	47.5%	47.5%	47.6%	47.7%	47.8%	47.8%	47.9%	48.0%	48.0%	48.1%	47.9%	47.8%
Workers with 1+ vehicles: Carpool	9.8%	9.7%	9.6%	9.5%	9.4%	9.3%	9.2%	9.1%	9.1%	9.0%	8.9%	8.8%	8.7%	8.7%	8.6%
Workers with 1+ vehicles: Transit	26.9%	26.7%	26.6%	26.5%	26.4%	26.3%	26.2%	26.1%	25.9%	25.8%	25.7%	25.6%	25.5%	25.8%	26.0%
Workers with 1+ vehicles: Taxi, walk, bike, other	9.6%	9.7%	9.7%	9.7%	9.8%	9.8%	9.9%	9.9%	9.9%	10.0%	10.0%	10.0%	10.1%	10.0%	10.0%
Workers with 1+ vehicles: Work at home	6.5%	6.6%	6.7%	6.8%	6.9%	7.0%	7.1%	7.2%	7.3%	7.4%	7.5%	7.6%	7.6%	7.6%	7.6%

Commute Mode Share Data

	Oct-2006	Nov-2006	Dec-2006	Jan-2007	Feb-2007	Mar-2007	Apr-2007	May-2007	Jun-2007	Jul-2007	Aug-2007	Sep-2007	Oct-2007	Nov-2007	Dec-2007
Commute Mode Shares															
Drive-Alone	40.0%	39.9%	39.7%	39.6%	39.4%	39.3%	39.1%	39.0%	38.8%	38.7%	38.6%	38.6%	38.6%	38.6%	38.5%
Carpool	7.5%	7.5%	7.4%	7.4%	7.3%	7.3%	7.2%	7.2%	7.1%	7.1%	7.2%	7.3%	7.4%	7.5%	7.6%
Transit	31.0%	31.2%	31.4%	31.7%	31.9%	32.1%	32.3%	32.5%	32.7%	33.0%	32.9%	32.8%	32.7%	32.6%	32.5%
Walk	9.6%	9.6%	9.6%	9.6%	9.6%	9.6%	9.7%	9.7%	9.7%	9.7%	9.6%	9.6%	9.6%	9.6%	9.5%
Taxi, bike, other	4.5%	4.5%	4.6%	4.6%	4.7%	4.7%	4.7%	4.8%	4.8%	4.9%	4.8%	4.8%	4.8%	4.8%	4.7%
Work at home	7.3%	7.3%	7.2%	7.2%	7.1%	7.0%	7.0%	6.9%	6.8%	6.8%	6.8%	6.9%	7.0%	7.0%	7.1%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	34.7%	34.5%	34.3%	34.1%	34.0%	33.8%	33.6%	33.4%	33.2%	33.0%	33.0%	33.0%	33.0%	33.0%	33.0%
Workers earning $0-50k: Carpool	7.9%	7.8%	7.6%	7.5%	7.4%	7.3%	7.2%	7.1%	7.0%	6.9%	7.0%	7.2%	7.3%	7.4%	7.6%
Workers earning $0-50k: Transit	34.7%	35.0%	35.2%	35.5%	35.7%	36.0%	36.2%	36.5%	36.7%	37.0%	36.8%	36.7%	36.6%	36.4%	36.3%
Workers earning $0-50k: Taxi, walk, bike, other	15.6%	15.6%	15.6%	15.6%	15.5%	15.5%	15.5%	15.5%	15.5%	15.5%	15.5%	15.6%	15.7%	15.8%	15.8%
Workers earning $0-50k: Work at home	7.1%	7.2%	7.2%	7.3%	7.3%	7.4%	7.5%	7.5%	7.6%	7.6%	7.6%	7.5%	7.4%	7.4%	7.3%
Workers earning $50k+: Drive-Alone	46.4%	46.3%	46.1%	46.0%	45.8%	45.7%	45.6%	45.4%	45.3%	45.2%	45.1%	45.0%	44.9%	44.8%	44.7%
Workers earning $50k+: Carpool	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.3%	7.4%	7.5%	7.6%	7.7%
Workers earning $50k+: Transit	26.4%	26.7%	26.9%	27.1%	27.3%	27.5%	27.7%	28.0%	28.2%	28.4%	28.3%	28.3%	28.3%	28.2%	28.2%
Workers earning $50k+: Taxi, walk, bike, other	12.3%	12.4%	12.6%	12.7%	12.8%	13.0%	13.1%	13.2%	13.3%	13.4%	13.3%	13.1%	12.9%	12.7%	12.5%
Workers earning $50k+: Work at home	7.6%	7.4%	7.2%	7.0%	6.8%	6.6%	6.4%	6.2%	6.0%	5.8%	6.0%	6.2%	6.4%	6.6%	6.8%
Workers with 0 vehicles: Drive-Alone	7.3%	7.4%	7.4%	7.4%	7.5%	7.5%	7.5%	7.6%	7.6%	7.6%	7.6%	7.6%	7.6%	7.5%	7.5%
Workers with 0 vehicles: Carpool	3.0%	2.9%	2.8%	2.6%	2.5%	2.4%	2.3%	2.1%	2.0%	1.9%	2.0%	2.1%	2.1%	2.2%	2.2%
Workers with 0 vehicles: Transit	52.2%	52.3%	52.3%	52.4%	52.4%	52.5%	52.5%	52.5%	52.6%	52.6%	52.6%	52.6%	52.5%	52.5%	52.5%
Workers with 0 vehicles: Taxi, walk, bike, other	31.8%	32.0%	32.2%	32.4%	32.6%	32.8%	33.0%	33.2%	33.4%	33.6%	33.3%	32.9%	32.6%	32.3%	31.9%
Workers with 0 vehicles: Work at home	5.7%	5.5%	5.3%	5.2%	5.0%	4.8%	4.7%	4.5%	4.3%	4.2%	4.5%	4.9%	5.2%	5.5%	5.8%
Workers with 1+ vehicles: Drive-Alone	47.7%	47.5%	47.4%	47.3%	47.1%	47.0%	46.8%	46.7%	46.6%	46.4%	46.4%	46.4%	46.3%	46.3%	46.2%
Workers with 1+ vehicles: Carpool	8.6%	8.5%	8.5%	8.5%	8.4%	8.4%	8.4%	8.3%	8.3%	8.2%	8.4%	8.5%	8.6%	8.8%	8.9%
Workers with 1+ vehicles: Transit	26.3%	26.5%	26.7%	27.0%	27.2%	27.5%	27.7%	28.0%	28.2%	28.4%	28.3%	28.2%	28.0%	27.9%	27.8%
Workers with 1+ vehicles: Taxi, walk, bike, other	10.0%	9.9%	9.9%	9.9%	9.8%	9.8%	9.8%	9.8%	9.7%	9.7%	9.7%	9.8%	9.8%	9.8%	9.9%
Workers with 1+ vehicles: Work at home	7.5%	7.5%	7.4%	7.4%	7.4%	7.3%	7.3%	7.3%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.3%

Commute Mode Share Data

	Jan-2008	Feb-2008	Mar-2008	Apr-2008	May-2008	Jun-2008	Jul-2008	Aug-2008	Sep-2008	Oct-2008	Nov-2008	Dec-2008	Jan-2009	Feb-2009	Mar-2009
Commute Mode Shares															
Drive-Alone	38.5%	38.5%	38.5%	38.4%	38.4%	38.4%	38.4%	38.4%	38.5%	38.5%	38.6%	38.6%	38.6%	38.7%	38.7%
Carpool	7.7%	7.8%	7.9%	8.0%	8.1%	8.3%	8.4%	8.3%	8.2%	8.1%	8.1%	8.0%	7.9%	7.8%	7.8%
Transit	32.4%	32.3%	32.2%	32.1%	32.0%	31.9%	31.9%	31.9%	31.8%	31.8%	31.8%	31.8%	31.8%	31.8%	31.8%
Walk	9.5%	9.5%	9.5%	9.5%	9.4%	9.4%	9.4%	9.5%	9.6%	9.6%	9.7%	9.8%	9.9%	10.0%	10.0%
Taxi, bike, other	4.7%	4.7%	4.6%	4.6%	4.6%	4.6%	4.5%	4.6%	4.6%	4.6%	4.6%	4.6%	4.7%	4.7%	4.7%
Work at home	7.1%	7.2%	7.3%	7.3%	7.4%	7.4%	7.5%	7.4%	7.4%	7.3%	7.2%	7.2%	7.1%	7.0%	7.0%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	33.0%	33.0%	33.0%	33.0%	33.0%	33.0%	33.0%	33.1%	33.3%	33.5%	33.6%	33.8%	33.9%	34.1%	34.3%
Workers earning $0-50k: Carpool	7.7%	7.8%	8.0%	8.1%	8.2%	8.3%	8.5%	8.3%	8.2%	8.0%	7.9%	7.7%	7.6%	7.4%	7.3%
Workers earning $0-50k: Transit	36.2%	36.0%	35.9%	35.8%	35.7%	35.5%	35.4%	35.3%	35.2%	35.1%	35.1%	35.0%	34.9%	34.8%	34.7%
Workers earning $0-50k: Taxi, walk, bike, other	15.9%	16.0%	16.0%	16.1%	16.2%	16.2%	16.3%	16.4%	16.4%	16.5%	16.6%	16.6%	16.7%	16.8%	16.8%
Workers earning $0-50k: Work at home	7.2%	7.2%	7.1%	7.1%	7.0%	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%	6.9%
Workers earning $50k+: Drive-Alone	44.7%	44.6%	44.5%	44.4%	44.4%	44.3%	44.2%	44.2%	44.1%	44.1%	44.1%	44.0%	44.0%	43.9%	43.9%
Workers earning $50k+: Carpool	7.8%	7.8%	7.9%	8.0%	8.1%	8.2%	8.2%	8.2%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%	8.3%
Workers earning $50k+: Transit	28.2%	28.1%	28.1%	28.1%	28.1%	28.0%	28.0%	28.1%	28.1%	28.2%	28.2%	28.3%	28.3%	28.4%	28.4%
Workers earning $50k+: Taxi, walk, bike, other	12.4%	12.2%	12.0%	11.9%	11.7%	11.5%	11.4%	11.5%	11.6%	11.7%	11.8%	11.9%	12.1%	12.2%	12.3%
Workers earning $50k+: Work at home	7.0%	7.2%	7.4%	7.6%	7.8%	8.0%	8.2%	8.0%	7.9%	7.7%	7.6%	7.5%	7.3%	7.2%	7.0%
Workers with 0 vehicles: Drive-Alone	7.5%	7.4%	7.4%	7.4%	7.4%	7.3%	7.3%	7.3%	7.2%	7.1%	7.1%	7.0%	6.9%	6.9%	6.8%
Workers with 0 vehicles: Carpool	2.3%	2.4%	2.4%	2.5%	2.5%	2.6%	2.7%	2.7%	2.7%	2.8%	2.8%	2.9%	2.9%	3.0%	3.0%
Workers with 0 vehicles: Transit	52.5%	52.4%	52.4%	52.4%	52.4%	52.3%	52.3%	52.3%	52.2%	52.2%	52.1%	52.1%	52.1%	52.0%	52.0%
Workers with 0 vehicles: Taxi, walk, bike, other	31.6%	31.3%	31.0%	30.6%	30.3%	30.0%	29.7%	29.9%	30.1%	30.3%	30.5%	30.7%	30.9%	31.1%	31.3%
Workers with 0 vehicles: Work at home	6.2%	6.5%	6.8%	7.1%	7.4%	7.7%	8.0%	7.9%	7.8%	7.6%	7.5%	7.3%	7.2%	7.1%	6.9%
Workers with 1+ vehicles: Drive-Alone	46.2%	46.1%	46.1%	46.1%	46.0%	46.0%	45.9%	46.0%	46.1%	46.2%	46.3%	46.4%	46.5%	46.6%	46.7%
Workers with 1+ vehicles: Carpool	9.0%	9.1%	9.2%	9.4%	9.5%	9.6%	9.7%	9.6%	9.5%	9.4%	9.3%	9.2%	9.1%	9.0%	8.9%
Workers with 1+ vehicles: Transit	27.6%	27.5%	27.4%	27.3%	27.1%	27.0%	26.9%	26.9%	26.9%	26.9%	26.8%	26.8%	26.8%	26.8%	26.8%
Workers with 1+ vehicles: Taxi, walk, bike, other	9.9%	9.9%	10.0%	10.0%	10.0%	10.1%	10.1%	10.2%	10.2%	10.3%	10.3%	10.4%	10.5%	10.5%	10.6%
Workers with 1+ vehicles: Work at home	7.3%	7.3%	7.3%	7.3%	7.3%	7.4%	7.4%	7.3%	7.3%	7.2%	7.2%	7.1%	7.1%	7.0%	7.0%

Commute Mode Share Data

	Apr-2009	May-2009	Jun-2009	Jul-2009	Aug-2009	Sep-2009	Oct-2009	Nov-2009	Dec-2009	Jan-2010	Feb-2010	Mar-2010	Apr-2010	May-2010	Jun-2010
Commute Mode Shares															
Drive-Alone	38.8%	38.8%	38.9%	38.9%	38.7%	38.5%	38.2%	38.0%	37.7%	37.5%	37.2%	37.0%	36.8%	36.5%	36.3%
Carpool	7.7%	7.6%	7.5%	7.5%	7.5%	7.5%	7.6%	7.6%	7.6%	7.7%	7.7%	7.8%	7.8%	7.8%	7.9%
Transit	31.8%	31.8%	31.8%	31.8%	32.0%	32.2%	32.4%	32.5%	32.7%	32.9%	33.1%	33.3%	33.5%	33.7%	33.9%
Walk	10.1%	10.2%	10.3%	10.4%	10.3%	10.2%	10.1%	10.1%	10.0%	9.9%	9.8%	9.7%	9.7%	9.6%	9.5%
Taxi, bike, other	4.7%	4.7%	4.7%	4.8%	4.9%	4.9%	5.0%	5.1%	5.2%	5.3%	5.4%	5.5%	5.6%	5.7%	5.8%
Work at home	6.9%	6.8%	6.8%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	34.4%	34.6%	34.7%	34.9%	34.6%	34.3%	34.0%	33.7%	33.4%	33.1%	32.8%	32.5%	32.2%	31.9%	31.6%
Workers earning $0-50k: Carpool	7.1%	7.0%	6.8%	6.7%	6.8%	6.9%	7.0%	7.1%	7.3%	7.4%	7.5%	7.6%	7.7%	7.9%	8.0%
Workers earning $0-50k: Transit	34.6%	34.6%	34.5%	34.4%	34.6%	34.8%	34.9%	35.1%	35.3%	35.5%	35.7%	35.9%	36.1%	36.3%	36.4%
Workers earning $0-50k: Taxi, walk, bike, other	16.9%	16.9%	17.0%	17.1%	17.0%	17.0%	17.0%	17.0%	16.9%	16.9%	16.9%	16.9%	16.8%	16.8%	16.8%
Workers earning $0-50k: Work at home	6.9%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%	7.1%	7.1%	7.1%	7.1%	7.2%	7.2%	7.2%
Workers earning $50k+: Drive-Alone	43.9%	43.8%	43.8%	43.7%	43.5%	43.4%	43.2%	43.0%	42.8%	42.6%	42.4%	42.2%	42.0%	41.8%	41.6%
Workers earning $50k+: Carpool	8.3%	8.4%	8.4%	8.4%	8.3%	8.3%	8.2%	8.1%	8.1%	8.0%	8.0%	7.9%	7.9%	7.8%	7.8%
Workers earning $50k+: Transit	28.5%	28.6%	28.6%	28.7%	28.9%	29.1%	29.3%	29.5%	29.7%	29.9%	30.1%	30.3%	30.5%	30.7%	30.9%
Workers earning $50k+: Taxi, walk, bike, other	12.4%	12.5%	12.7%	12.8%	12.9%	12.9%	13.0%	13.1%	13.2%	13.2%	13.3%	13.4%	13.5%	13.5%	13.6%
Workers earning $50k+: Work at home	6.9%	6.7%	6.6%	6.4%	6.4%	6.4%	6.3%	6.3%	6.3%	6.2%	6.2%	6.2%	6.1%	6.1%	6.1%
Workers with 0 vehicles: Drive-Alone	6.7%	6.7%	6.6%	6.6%	6.8%	7.1%	7.4%	7.7%	7.9%	8.2%	8.4%	8.7%	8.9%	9.2%	9.4%
Workers with 0 vehicles: Carpool	3.1%	3.1%	3.1%	3.2%	3.1%	3.1%	3.1%	3.0%	3.0%	2.9%	2.9%	2.9%	2.8%	2.8%	2.8%
Workers with 0 vehicles: Transit	51.9%	51.9%	51.9%	51.8%	52.1%	52.4%	52.7%	52.9%	53.2%	53.5%	53.7%	54.0%	54.2%	54.5%	54.7%
Workers with 0 vehicles: Taxi, walk, bike, other	31.5%	31.7%	31.9%	32.1%	31.6%	31.1%	30.6%	30.2%	29.7%	29.3%	28.8%	28.4%	28.0%	27.6%	27.1%
Workers with 0 vehicles: Work at home	6.8%	6.7%	6.5%	6.4%	6.3%	6.3%	6.3%	6.2%	6.2%	6.1%	6.1%	6.1%	6.0%	6.0%	5.9%
Workers with 1+ vehicles: Drive-Alone	46.8%	47.0%	47.1%	47.2%	46.9%	46.5%	46.2%	45.9%	45.6%	45.3%	45.0%	44.6%	44.3%	44.0%	43.7%
Workers with 1+ vehicles: Carpool	8.8%	8.6%	8.5%	8.4%	8.5%	8.6%	8.7%	8.7%	8.8%	8.9%	9.0%	9.0%	9.1%	9.2%	9.3%
Workers with 1+ vehicles: Transit	26.8%	26.8%	26.8%	26.8%	26.9%	27.0%	27.1%	27.2%	27.4%	27.5%	27.6%	27.7%	27.9%	28.0%	28.1%
Workers with 1+ vehicles: Taxi, walk, bike, other	10.7%	10.7%	10.8%	10.9%	11.0%	11.1%	11.2%	11.3%	11.4%	11.5%	11.6%	11.8%	11.9%	12.0%	12.1%
Workers with 1+ vehicles: Work at home	6.9%	6.9%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%

Commute Mode Share Data

	Jul-2010	Aug-2010	Sep-2010	Oct-2010	Nov-2010	Dec-2010	Jan-2011	Feb-2011	Mar-2011	Apr-2011	May-2011	Jun-2011	Jul-2011	Aug-2011	Sep-2011
Commute Mode Shares															
Drive-Alone	36.0%	36.2%	36.3%	36.4%	36.6%	36.7%	36.8%	37.0%	37.1%	37.2%	37.4%	37.5%	37.7%	37.5%	37.4%
Carpool	7.9%	7.9%	7.8%	7.8%	7.7%	7.7%	7.6%	7.6%	7.5%	7.5%	7.4%	7.4%	7.3%	7.3%	7.4%
Transit	34.1%	33.8%	33.6%	33.4%	33.2%	33.0%	32.8%	32.6%	32.4%	32.2%	32.0%	31.8%	31.6%	31.7%	31.8%
Walk	9.4%	9.5%	9.5%	9.6%	9.6%	9.6%	9.7%	9.7%	9.7%	9.8%	9.8%	9.8%	9.9%	9.9%	9.9%
Taxi, bike, other	5.9%	5.9%	5.9%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.8%	5.9%
Work at home	6.7%	6.8%	6.9%	7.0%	7.1%	7.2%	7.3%	7.4%	7.5%	7.5%	7.6%	7.7%	7.8%	7.7%	7.7%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	31.3%	31.5%	31.7%	31.9%	32.1%	32.3%	32.5%	32.7%	32.8%	33.0%	33.2%	33.4%	33.6%	33.5%	33.4%
Workers earning $0-50k: Carpool	8.1%	8.1%	8.0%	7.9%	7.9%	7.8%	7.7%	7.7%	7.6%	7.6%	7.5%	7.4%	7.4%	7.4%	7.3%
Workers earning $0-50k: Transit	36.6%	36.3%	36.1%	35.8%	35.5%	35.2%	34.9%	34.7%	34.4%	34.1%	33.8%	33.5%	33.2%	33.5%	33.8%
Workers earning $0-50k: Taxi, walk, bike, other	16.8%	16.8%	16.8%	16.8%	16.9%	16.9%	16.9%	17.0%	17.0%	17.0%	17.1%	17.1%	17.1%	17.1%	17.0%
Workers earning $0-50k: Work at home	7.2%	7.3%	7.5%	7.6%	7.7%	7.8%	7.9%	8.0%	8.2%	8.3%	8.4%	8.5%	8.6%	8.5%	8.4%
Workers earning $50k+: Drive-Alone	41.4%	41.5%	41.5%	41.6%	41.7%	41.7%	41.8%	41.9%	41.9%	42.0%	42.1%	42.1%	42.2%	42.0%	41.9%
Workers earning $50k+: Carpool	7.7%	7.7%	7.6%	7.6%	7.5%	7.5%	7.5%	7.4%	7.4%	7.4%	7.3%	7.3%	7.3%	7.3%	7.4%
Workers earning $50k+: Transit	31.1%	31.0%	30.9%	30.8%	30.6%	30.5%	30.4%	30.3%	30.1%	30.0%	29.9%	29.8%	29.6%	29.6%	29.7%
Workers earning $50k+: Taxi, walk, bike, other	13.7%	13.7%	13.7%	13.8%	13.8%	13.8%	13.8%	13.8%	13.9%	13.9%	13.9%	13.9%	13.9%	14.1%	14.3%
Workers earning $50k+: Work at home	6.1%	6.1%	6.2%	6.3%	6.4%	6.4%	6.5%	6.6%	6.7%	6.7%	6.8%	6.9%	7.0%	6.9%	6.8%
Workers with 0 vehicles: Drive-Alone	9.7%	9.5%	9.3%	9.1%	8.9%	8.7%	8.5%	8.3%	8.1%	7.9%	7.6%	7.4%	7.2%	7.3%	7.4%
Workers with 0 vehicles: Carpool	2.7%	2.8%	2.8%	2.8%	2.9%	2.9%	3.0%	3.0%	3.1%	3.1%	3.1%	3.2%	3.2%	3.1%	2.9%
Workers with 0 vehicles: Transit	55.0%	54.8%	54.7%	54.5%	54.4%	54.2%	54.0%	53.9%	53.7%	53.6%	53.4%	53.2%	53.1%	53.0%	53.0%
Workers with 0 vehicles: Taxi, walk, bike, other	26.7%	27.1%	27.4%	27.8%	28.1%	28.5%	28.9%	29.2%	29.6%	30.0%	30.4%	30.8%	31.2%	31.3%	31.3%
Workers with 0 vehicles: Work at home	5.9%	5.9%	5.8%	5.8%	5.7%	5.7%	5.6%	5.6%	5.5%	5.5%	5.5%	5.4%	5.4%	5.3%	5.3%
Workers with 1+ vehicles: Drive-Alone	43.4%	43.6%	43.7%	43.9%	44.1%	44.3%	44.5%	44.7%	44.8%	45.0%	45.2%	45.4%	45.6%	45.5%	45.3%
Workers with 1+ vehicles: Carpool	9.4%	9.3%	9.2%	9.1%	9.0%	8.9%	8.9%	8.8%	8.7%	8.6%	8.5%	8.5%	8.4%	8.5%	8.6%
Workers with 1+ vehicles: Transit	28.2%	28.0%	27.9%	27.7%	27.5%	27.3%	27.2%	27.0%	26.8%	26.6%	26.4%	26.3%	26.1%	26.2%	26.3%
Workers with 1+ vehicles: Taxi, walk, bike, other	12.2%	12.2%	12.1%	12.1%	12.1%	12.0%	12.0%	12.0%	11.9%	11.9%	11.9%	11.8%	11.8%	11.8%	11.8%
Workers with 1+ vehicles: Work at home	6.8%	7.0%	7.1%	7.2%	7.3%	7.4%	7.5%	7.6%	7.7%	7.8%	7.9%	8.0%	8.1%	8.0%	7.9%

Commute Mode Share Data

	Oct-2011	Nov-2011	Dec-2011	Jan-2012	Feb-2012	Mar-2012	Apr-2012	May-2012	Jun-2012	Jul-2012	Aug-2012	Sep-2012	Oct-2012	Nov-2012	Dec-2012
Commute Mode Shares															
Drive-Alone	37.3%	37.2%	37.1%	37.0%	36.9%	36.8%	36.6%	36.5%	36.4%	36.3%	36.3%	36.3%	36.3%	36.3%	36.4%
Carpool	7.4%	7.4%	7.5%	7.5%	7.5%	7.6%	7.6%	7.6%	7.7%	7.7%	7.6%	7.5%	7.5%	7.4%	7.3%
Transit	32.0%	32.1%	32.2%	32.3%	32.5%	32.6%	32.7%	32.9%	33.0%	33.1%	33.1%	33.0%	33.0%	33.0%	32.9%
Walk	9.9%	9.9%	9.9%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.9%	10.0%	10.1%	10.2%	10.3%
Taxi, bike, other	5.9%	6.0%	6.0%	6.1%	6.1%	6.1%	6.2%	6.2%	6.3%	6.3%	6.3%	6.4%	6.4%	6.4%	6.4%
Work at home	7.6%	7.5%	7.4%	7.3%	7.2%	7.1%	7.0%	6.9%	6.8%	6.7%	6.7%	6.7%	6.7%	6.7%	6.7%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	33.3%	33.2%	33.1%	33.0%	32.9%	32.8%	32.6%	32.5%	32.4%	32.3%	32.4%	32.4%	32.5%	32.5%	32.6%
Workers earning $0-50k: Carpool	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%
Workers earning $0-50k: Transit	34.0%	34.3%	34.6%	34.9%	35.1%	35.4%	35.7%	36.0%	36.3%	36.5%	36.4%	36.2%	36.0%	35.9%	35.7%
Workers earning $0-50k: Taxi, walk, bike, other	17.0%	17.0%	16.9%	16.9%	16.8%	16.8%	16.7%	16.7%	16.6%	16.6%	16.8%	17.0%	17.1%	17.3%	17.5%
Workers earning $0-50k: Work at home	8.3%	8.2%	8.1%	8.0%	7.9%	7.8%	7.6%	7.5%	7.4%	7.3%	7.2%	7.2%	7.1%	7.0%	7.0%
Workers earning $50k+: Drive-Alone	41.7%	41.5%	41.4%	41.2%	41.1%	40.9%	40.7%	40.6%	40.4%	40.3%	40.2%	40.2%	40.1%	40.0%	40.0%
Workers earning $50k+: Carpool	7.5%	7.6%	7.6%	7.7%	7.8%	7.9%	7.9%	8.0%	8.1%	8.1%	8.0%	7.8%	7.6%	7.5%	7.3%
Workers earning $50k+: Transit	29.7%	29.7%	29.7%	29.7%	29.7%	29.7%	29.7%	29.7%	29.7%	29.7%	29.8%	29.9%	30.0%	30.1%	30.2%
Workers earning $50k+: Taxi, walk, bike, other	14.4%	14.6%	14.7%	14.9%	15.0%	15.2%	15.3%	15.4%	15.6%	15.7%	15.7%	15.8%	15.8%	15.9%	15.9%
Workers earning $50k+: Work at home	6.7%	6.7%	6.6%	6.5%	6.5%	6.4%	6.3%	6.3%	6.2%	6.2%	6.2%	6.3%	6.4%	6.5%	6.5%
Workers with 0 vehicles: Drive-Alone	7.6%	7.7%	7.8%	7.9%	8.0%	8.1%	8.2%	8.3%	8.4%	8.6%	8.4%	8.2%	8.0%	7.9%	7.7%
Workers with 0 vehicles: Carpool	2.7%	2.5%	2.4%	2.2%	2.1%	1.9%	1.7%	1.6%	1.4%	1.3%	1.3%	1.3%	1.3%	1.3%	1.4%
Workers with 0 vehicles: Transit	53.0%	53.0%	52.9%	52.9%	52.9%	52.9%	52.8%	52.8%	52.8%	52.8%	52.9%	53.0%	53.1%	53.3%	53.4%
Workers with 0 vehicles: Taxi, walk, bike, other	31.4%	31.5%	31.6%	31.7%	31.8%	31.8%	31.9%	32.0%	32.1%	32.1%	32.2%	32.3%	32.4%	32.5%	32.6%
Workers with 0 vehicles: Work at home	5.3%	5.3%	5.3%	5.3%	5.3%	5.3%	5.3%	5.3%	5.2%	5.2%	5.2%	5.1%	5.1%	5.0%	5.0%
Workers with 1+ vehicles: Drive-Alone	45.2%	45.1%	45.0%	44.8%	44.7%	44.6%	44.5%	44.3%	44.2%	44.1%	44.1%	44.0%	44.0%	43.9%	43.9%
Workers with 1+ vehicles: Carpool	8.7%	8.7%	8.8%	8.9%	9.0%	9.1%	9.2%	9.3%	9.4%	9.5%	9.3%	9.2%	9.1%	9.0%	8.8%
Workers with 1+ vehicles: Transit	26.5%	26.6%	26.7%	26.9%	27.0%	27.1%	27.2%	27.4%	27.5%	27.6%	27.6%	27.6%	27.7%	27.7%	27.7%
Workers with 1+ vehicles: Taxi, walk, bike, other	11.8%	11.8%	11.8%	11.8%	11.8%	11.8%	11.8%	11.8%	11.8%	11.8%	12.0%	12.1%	12.3%	12.4%	12.5%
Workers with 1+ vehicles: Work at home	7.8%	7.8%	7.7%	7.6%	7.5%	7.4%	7.3%	7.2%	7.1%	7.0%	7.0%	7.0%	7.0%	7.0%	7.1%

Commute Mode Share Data

	Jan-2013	Feb-2013	Mar-2013	Apr-2013	May-2013	Jun-2013	Jul-2013	Aug-2013	Sep-2013	Oct-2013	Nov-2013	Dec-2013	Jan-2014	Feb-2014	Mar-2014
Commute Mode Shares															
Drive-Alone	36.4%	36.4%	36.4%	36.4%	36.4%	36.4%	36.4%	36.2%	36.0%	35.9%	35.7%	35.5%	35.3%	35.1%	34.9%
Carpool	7.2%	7.1%	7.1%	7.0%	6.9%	6.8%	6.8%	6.8%	6.8%	6.9%	6.9%	6.9%	6.9%	7.0%	7.0%
Transit	32.9%	32.9%	32.8%	32.8%	32.7%	32.7%	32.7%	32.8%	32.9%	33.0%	33.1%	33.2%	33.3%	33.5%	33.6%
Walk	10.4%	10.5%	10.6%	10.7%	10.7%	10.8%	10.9%	10.9%	11.0%	11.0%	11.0%	11.0%	11.1%	11.1%	11.1%
Taxi, bike, other	6.4%	6.4%	6.4%	6.4%	6.4%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%
Work at home	6.7%	6.7%	6.7%	6.7%	6.7%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.9%	6.9%	6.9%
Commute Mode Shares by Segment															
Workers earning $0-50k: Drive-Alone	32.6%	32.7%	32.7%	32.8%	32.9%	32.9%	33.0%	32.8%	32.6%	32.5%	32.3%	32.2%	32.0%	31.8%	31.7%
Workers earning $0-50k: Carpool	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.3%	7.2%	7.2%	7.2%	7.2%	7.2%
Workers earning $0-50k: Transit	35.6%	35.4%	35.2%	35.1%	34.9%	34.7%	34.6%	34.7%	34.8%	35.0%	35.1%	35.2%	35.3%	35.5%	35.6%
Workers earning $0-50k: Taxi, walk, bike, other	17.7%	17.8%	18.0%	18.2%	18.4%	18.6%	18.7%	18.7%	18.7%	18.7%	18.6%	18.6%	18.6%	18.6%	18.6%
Workers earning $0-50k: Work at home	6.9%	6.8%	6.7%	6.7%	6.6%	6.5%	6.5%	6.5%	6.6%	6.6%	6.7%	6.8%	6.8%	6.9%	6.9%
Workers earning $50k+: Drive-Alone	39.9%	39.9%	39.8%	39.7%	39.7%	39.6%	39.6%	39.4%	39.1%	38.9%	38.7%	38.5%	38.3%	38.1%	37.9%
Workers earning $50k+: Carpool	7.2%	7.0%	6.9%	6.7%	6.6%	6.5%	6.3%	6.4%	6.4%	6.5%	6.5%	6.6%	6.7%	6.7%	6.8%
Workers earning $50k+: Transit	30.3%	30.5%	30.6%	30.7%	30.7%	30.8%	30.9%	31.0%	31.1%	31.2%	31.3%	31.4%	31.5%	31.6%	31.7%
Workers earning $50k+: Taxi, walk, bike, other	15.9%	16.0%	16.0%	16.1%	16.1%	16.1%	16.2%	16.2%	16.3%	16.4%	16.5%	16.5%	16.6%	16.7%	16.8%
Workers earning $50k+: Work at home	6.6%	6.7%	6.7%	6.8%	6.9%	6.9%	7.0%	7.0%	7.0%	7.0%	6.9%	6.9%	6.9%	6.9%	6.9%
Workers with 0 vehicles: Drive-Alone	7.5%	7.3%	7.2%	7.0%	6.8%	6.6%	6.4%	6.2%	6.0%	5.8%	5.7%	5.5%	5.3%	5.2%	5.0%
Workers with 0 vehicles: Carpool	1.4%	1.4%	1.4%	1.4%	1.4%	1.4%	1.5%	1.6%	1.7%	1.8%	1.9%	2.0%	2.2%	2.3%	2.4%
Workers with 0 vehicles: Transit	53.5%	53.6%	53.8%	53.9%	54.0%	54.2%	54.3%	54.3%	54.2%	54.2%	54.1%	54.0%	54.0%	53.9%	53.9%
Workers with 0 vehicles: Taxi, walk, bike, other	32.7%	32.8%	32.9%	33.0%	33.1%	33.2%	33.3%	33.3%	33.3%	33.4%	33.4%	33.4%	33.4%	33.5%	33.5%
Workers with 0 vehicles: Work at home	4.9%	4.8%	4.8%	4.7%	4.7%	4.6%	4.5%	4.6%	4.7%	4.8%	4.9%	5.0%	5.1%	5.2%	5.3%
Workers with 1+ vehicles: Drive-Alone	43.8%	43.8%	43.8%	43.7%	43.7%	43.6%	43.6%	43.5%	43.4%	43.3%	43.2%	43.1%	43.0%	42.9%	42.7%
Workers with 1+ vehicles: Carpool	8.7%	8.6%	8.5%	8.3%	8.2%	8.1%	8.0%	8.0%	8.0%	8.1%	8.1%	8.1%	8.1%	8.2%	8.2%
Workers with 1+ vehicles: Transit	27.7%	27.7%	27.7%	27.7%	27.8%	27.8%	27.8%	27.9%	27.9%	28.0%	28.1%	28.2%	28.3%	28.4%	28.5%
Workers with 1+ vehicles: Taxi, walk, bike, other	12.7%	12.8%	12.9%	13.1%	13.2%	13.3%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%
Workers with 1+ vehicles: Work at home	7.1%	7.1%	7.1%	7.1%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%

Commute Mode Share Data

	Apr-2014	May-2014	Jun-2014	Jul-2014	Aug-2014	Sep-2014	Oct-2014	Nov-2014	Dec-2014
Commute Mode Shares									
Drive-Alone	34.8%	34.6%	34.4%	34.2%	34.0%	33.9%	33.7%	33.5%	33.3%
Carpool	7.0%	7.0%	7.1%	7.1%	7.1%	7.2%	7.2%	7.2%	7.2%
Transit	33.7%	33.8%	33.9%	34.0%	34.1%	34.2%	34.3%	34.4%	34.5%
Walk	11.1%	11.2%	11.2%	11.2%	11.2%	11.3%	11.3%	11.3%	11.3%
Taxi, bike, other	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%	6.5%
Work at home	6.9%	6.9%	6.9%	7.0%	7.0%	7.0%	7.0%	7.0%	7.0%
Commute Mode Shares by Segment									
Workers earning $0-50k: Drive-Alone	31.5%	31.4%	31.2%	31.0%	30.9%	30.7%	30.5%	30.4%	30.2%
Workers earning $0-50k: Carpool	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%	7.2%
Workers earning $0-50k: Transit	35.7%	35.9%	36.0%	36.1%	36.2%	36.4%	36.5%	36.6%	36.8%
Workers earning $0-50k: Taxi, walk, bike, other	18.5%	18.5%	18.5%	18.5%	18.5%	18.4%	18.4%	18.4%	18.4%
Workers earning $0-50k: Work at home	7.0%	7.0%	7.1%	7.2%	7.2%	7.3%	7.3%	7.4%	7.4%
Workers earning $50k+: Drive-Alone	37.7%	37.5%	37.2%	37.0%	36.8%	36.6%	36.4%	36.2%	36.0%
Workers earning $50k+: Carpool	6.8%	6.9%	6.9%	7.0%	7.0%	7.1%	7.2%	7.2%	7.3%
Workers earning $50k+: Transit	31.8%	31.9%	32.0%	32.1%	32.2%	32.3%	32.4%	32.5%	32.6%
Workers earning $50k+: Taxi, walk, bike, other	16.8%	16.9%	17.0%	17.1%	17.1%	17.2%	17.3%	17.3%	17.4%
Workers earning $50k+: Work at home	6.8%	6.8%	6.8%	6.8%	6.8%	6.8%	6.7%	6.7%	6.7%
Workers with 0 vehicles: Drive-Alone	4.8%	4.7%	4.5%	4.4%	4.2%	4.1%	3.9%	3.8%	3.7%
Workers with 0 vehicles: Carpool	2.5%	2.6%	2.7%	2.8%	2.9%	3.0%	3.1%	3.2%	3.2%
Workers with 0 vehicles: Transit	53.8%	53.8%	53.7%	53.7%	53.6%	53.6%	53.5%	53.5%	53.5%
Workers with 0 vehicles: Taxi, walk, bike, other	33.5%	33.5%	33.5%	33.6%	33.6%	33.6%	33.6%	33.6%	33.6%
Workers with 0 vehicles: Work at home	5.4%	5.4%	5.5%	5.6%	5.7%	5.8%	5.8%	5.9%	6.0%
Workers with 1+ vehicles: Drive-Alone	42.6%	42.5%	42.4%	42.3%	42.2%	42.1%	42.0%	41.9%	41.8%
Workers with 1+ vehicles: Carpool	8.2%	8.2%	8.2%	8.3%	8.3%	8.3%	8.3%	8.3%	8.4%
Workers with 1+ vehicles: Transit	28.6%	28.6%	28.7%	28.8%	28.9%	29.0%	29.1%	29.2%	29.2%
Workers with 1+ vehicles: Taxi, walk, bike, other	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.4%	13.3%
Workers with 1+ vehicles: Work at home	7.2%	7.2%	7.2%	7.2%	7.2%	7.3%	7.3%	7.3%	7.3%

Appendix C

Model Application Formulas

This appendix shows the formulas used to apply the models estimated in Chapter 4. The formulas for applying the preferred model of each type are derived from the general formula for each model type. While only the RegARIMA models are used for the remainder of the analysis, it is useful to show the ARIMA and regression formulas as a means of building towards those.

ARIMA Models

Equation 4.7 shows the general equation for a seasonal ARIMA model, and serves as the starting point for these model application formulas.

MUNI

Table 4.10 shows the estimation results for MUNI riders (y_t), which follows an $ARIMA(1,1,1)(0,1,0)_{12}$ form. Substituting these results into the general equation, gives:

$$(1-\phi_1 B)(1-B)(1-B^{12})y_t = (1+\theta_1 B)e_t \tag{C.1}$$

y_t is isolated on the left hand side, in a series of steps:

$$(1-\phi_1 B)(1-B-B^{12}+B^{13})y_t = (1+\theta_1 B)e_t \tag{C.2}$$

$$\begin{aligned}&\big(1-(1+\phi_1)B+\phi_1 B^2-B^{12}+(1+\phi_1)B^{13}-\phi_1 B^{14}\big)y_t=\\&(1+\theta_1 B)e_t\end{aligned} \tag{C.3}$$

$$\begin{aligned}&\big(1-B^{12}-(1+\phi_1)(B-B^{13})+\phi_1(B^2-B^{14})\big)y_t=\\&(1+\theta_1 B)e_t\end{aligned} \tag{C.4}$$

$$\begin{aligned}y_t =&B^{12}y_t+(1+\phi_1)(B-B^{13})y_t\\&-\phi_1(B^2-B^{14})y_t+\theta_1 Be_t+e_t\end{aligned} \tag{C.5}$$

$$\begin{aligned}y_t =&y_{t-12}+(1+\phi_1)(y_{t-1}-y_{t-13})\\&-\phi_1(y_{t-2}-y_{t-14})+\theta_1 e_{t-1}+e_t\end{aligned} \tag{C.6}$$

Adding the model coefficients gives:

$$\begin{aligned}y_t =&y_{t-12}+1.4708(y_{t-1}-y_{t-13})\\&-0.4708(y_{t-2}-y_{t-14})-0.9296e_{t-1}+e_t\end{aligned} \tag{C.7}$$

Equation C.7 is the final equation for predicting MUNI ridership using the preferred ARIMA model.

The interpretation of the autoregressive terms is that the current value is equal to the value from 12 months earlier, plus 147% of the year-over-year change from one month earlier, minus 47% of the year-over-year change from two months earlier. That portion of the equation is continuing the current rate of change. It is acceptable that the first term ($1.4708(y_{t-1}-y_{t-13})$) is greater than one, because it is offset by the second term, the autoregressive terms as a whole sum to one. The moving average component of the model ($-0.9296e_{t-1}$) serves to dampen any unexpected trends. If e_{t-1} is positive, it is because the change in the previous month (($y_{t-1}-y_{t-13}$) was bigger than expected, so subtracting off the previous error brings the model back in line with the longer term trend. e_t is the error in the current period, which is assumed to be white noise.

BART

Table 4.11 shows the estimation results for BART riders, which follows an $ARIMA(2,1,0)(0,1,1)_{12}$ form. Following equivalent logic to above, the model application formula is:

$$(1-\phi_1 B-\phi_2 B^2)(1-B)(1-B^{12})y_t = (1+\Theta_1 B^{12})e_t \tag{C.8}$$

$$(1-\phi_1 B-\phi_2 B^2)(1-B-B^{12}+B^{13})y_t = (1+\Theta_1 B^{12})e_t \tag{C.9}$$

$$\big(1-B^{12}-(1+\phi_1)(B-B^{13})+(\phi_1-\phi_2)(B^2-B^{14})+\phi_2(B^3-B^{15})\big)y_t = (1+\Theta_1 B^{12})e_t \tag{C.10}$$

$$\begin{aligned} y_t =& B^{12}y_t+(1+\phi_1)(B-B^{13})y_t \\ &-(\phi_1-\phi_2)(B^2-B^{14})y_t-\phi_2(B^3-B^{15})y_t \\ &+\Theta_1 B^{12}e_t+e_t \end{aligned} \tag{C.11}$$

$$\begin{aligned} y_t =& y_{t-12}+(1+\phi_1)(y_{t-1}-y_{t-13}) \\ &-(\phi_1-\phi_2)(y_{t-2}-y_{t-14})-\phi_2(y_{t-3}-y_{t-15}) \\ &+\Theta_1 e_{t-12}+e_t \end{aligned} \tag{C.12}$$

Adding the model coefficients gives:

$$\begin{aligned} y_t =& y_{t-12}+0.4023(y_{t-1}-y_{t-13}) \\ &+0.1483(y_{t-2}-y_{t-14})+0.4494(y_{t-3}-y_{t-15}) \\ &-0.8631e_{t-12}+e_t \end{aligned} \tag{C.13}$$

Note that the net autoregressive coefficients again sum to 1, representing a weighted average of the year-over-year change across the previous three months. Equation C.13 is the final equation for predicting BART ridership using the preferred ARIMA model.

Regression Models

Equation 4.9 shows the general equation for a regression model. The application form depends on the degree of differencing. A model with a single, seasonal difference can be expressed as:

$$(1-B^{12})y_t = \beta(1-B^{12})X_t + c + e_t \tag{C.14}$$

Which translates to:

$$y_t = B^{12}y_t + \beta(1-B^{12})X_t + c + e_t \tag{C.15}$$

$$y_t = y_{t-12} + \beta(X_t - X_{t-12}) + c + e_t \tag{C.16}$$

This equation says that the current value is equal to the value from one year prior, plus the change change contributed by the change in the regressors, plus a constant and an error term. The constant represents a linear trend beyond what is attributable to changes in the descriptive variables.

The MUNI and BART models have been estimated on data that have been transformed with a second difference, with one of the differences being seasonal. Neither includes a constant, which would be a quadratic trend in a model of second differences. Therefore, the models can be expressed as:

$$(1-B)(1-B^{12})y_t = \beta(1-B)(1-B^{12})X_t + e_t \tag{C.17}$$

Isolating y_t on the left hand side, in a series of steps, gives:

$$(1-B-B^{12}+B^{13})y_t = \beta(1-B-B^{12}+B^{13})X_t + e_t \tag{C.18}$$

$$
\begin{aligned}
y_t =& B^{12} y_t + \beta(1 - B^{12}) X_t \\
& + \left[(B - B^{13}) y_t - \beta(B - B^{13}) X_t\right] + e_t
\end{aligned}
\tag{C.19}
$$

$$
\begin{aligned}
y_t =& y_{t-12} + \beta(X_t - X_{t-12}) \\
& + \left[(y_{t-1} - y_{t-13}) - \beta(X_{t-1} - X_{t-13})\right] + e_t
\end{aligned}
\tag{C.20}
$$

The first two terms indicate that the current value is equal to the value from 12 months prior, plus the regression coefficients times the year-over-year change in the X values. The term in brackets includes the difference between the actual year-over-year change observed one month prior minus the year-over-year change that would have been predicted purely by the regressors. It plays a role similar to the constant in a model of first differences: representing trends beyond what is attributable to changes in descriptive variables. Finally, there is an error term.

MUNI

When the coefficients from the MUNI estimation are included, and the rate is converted to the total, the application equation becomes:

$$
\begin{aligned}
y_t =\ & y_{t-12} \\
& + 8536 \times (WkdyServMiles1000_t - WkdyServMiles1000_{t-12}) \\
& - 4352 \times (WkdyRailMiles1000_t - WkdyRailMiles1000_{t-12}) \\
& + 63927 \times (Runspeed_t - Runspeed_{t-12}) \\
& + 2.201 \times (EmpSF_t - EmpSF_{t-12}) \\
& + \big[(y_{t-1} - y_{t-13}) \\
& - \big(8536 \times (WkdyServMiles1000_{t-1} - WkdyServMiles1000_{t-13}) \\
& - 4352 \times (WkdyRailMiles1000_{t-1} - WkdyRailMiles1000_{t-13}) \\
& + 63927 \times (Runspeed_{t-1} - Runspeed_{t-13}) \\
& + 2.201 \times (EmpSF_{t-1} - EmpSF_{t-13})\big)\big] \\
& + e_t
\end{aligned}
\tag{C.21}
$$

BART

When the coefficients from the BART estimation are included, and the rate is converted to the total, the application equation becomes:

$$
\begin{aligned}
y_t = {} & y_{t-12} \\
& + 7613 \times (NumStations_t - NumStations_{t-12}) \\
& + 0.1827 \times (Emp4Cty_t - Emp4Cty_{t-12}) \\
& - 23490 \times (CashFare_t - CashFare_{t-12}) \\
& - 19690 \times (StrikeDays_t - StrikeDays_{t-12}) \\
& + \big[(y_{t-1} - y_{t-13}) \\
& - \big(7613 \times (NumStations_t - NumStations_{t-12}) \\
& + 0.1827 \times (Emp4Cty_t - Emp4Cty_{t-12}) \\
& - 23490 \times (CashFare_t - CashFare_{t-12}) \\
& - 19690 \times (StrikeDays_t - StrikeDays_{t-12})\big)\big] \\
& + e_t
\end{aligned} \tag{C.22}
$$

Equation C.21 is the final equation for predicting MUNI ridership using the preferred regression model model. Equation C.22 is the final equation for predicting BART ridership using the preferred regression model model.

Regression Models with ARIMA Errors

Equation 4.10 shows the general equation for a regression model with ARIMA errors. This combines elements elements of the application equations discussed in the previous two sections.

MUNI

The regression errors from the MUNI model takes the form $ARIMA(0,1,1)(0,1,0)_{12}$.

For the MUNI model, the model is:

$$\begin{aligned} y_t &= \beta X_t + n_t \\ (1-B)(1-B^{12})n_t &= (1+\theta_1 B)e_t \end{aligned} \tag{C.23}$$

$$\begin{aligned} y_t &= \beta X_t + n_t \\ (1-B-B^{12}+B^{13})n_t &= (1+\theta_1 B)e_t \end{aligned} \tag{C.24}$$

$$\begin{aligned} y_t &= \beta X_t + n_t \\ n_t &= B^{12}n_t + (B-B^{13})n_t + \theta_1 B e_t + e_t \end{aligned} \tag{C.25}$$

$$\begin{aligned} y_t &= \beta X_t + n_t \\ n_t &= n_{t-12} + (n_{t-1} - n_{t-13}) + \theta_1 e_{t-1} + e_t \end{aligned} \tag{C.26}$$

Adding the estimated model coefficients from Table 4.15 gives:

$$\begin{aligned} y_t =& 7971 \times WkdyServMiles1000_t \\ & - 2777 \times WkdyRailMiles1000_t \\ & + 49853 \times Runspeed_t \\ & + 0.876 \times EmpSF_t \\ & + n_t \\ n_t =& n_{t-12} + (n_{t-1} - n_{t-13}) - 0.309 e_{t-1} + e_t \end{aligned} \tag{C.27}$$

Equation C.27 is the final equation for the application of the MUNI regression model with ARIMA errors.

BART

The BART model (regression with $ARIMA(0,1,2)(0,1,1)_{12}$ errors) is:

$$\begin{aligned} y_t &= \beta X_t + n_t \\ (1-B)(1-B^{12})n_t &= (1+\theta_1 B + \theta_2 B^2)(1+\Theta_1 B^{12})e_t \end{aligned} \tag{C.28}$$

$$\begin{aligned}
y_t &= \beta X_t + n_t \\
(1-B-B^{12}+B^{13})n_t &= \\
(1+\theta_1 B+\theta_2 B^2+\Theta_1 B^{12}&+\theta_1\Theta_1 B^{13}+\theta_2\Theta_1 B^{14})e_t
\end{aligned} \tag{C.29}$$

$$\begin{aligned}
y_t &= \beta X_t + n_t \\
(1-B^{12}-(B-B^{13}))n_t &= \\
(1+\theta_1 B+\theta_2 B^2+\Theta_1 B^{12}&+\theta_1\Theta_1 B^{13}+\theta_2\Theta_1 B^{14})e_t
\end{aligned} \tag{C.30}$$

$$\begin{aligned}
y_t =&\beta X_t + n_t \\
n_t =&B^{12}n_t + (B-B^{13})n_t \\
&+\theta_1 Be_t+\theta_2 B^2 e_t+\Theta_1 B^{12}e_t+\theta_1\Theta_1 B^{13}e_t+\theta_2\Theta_1 B^{14}e_t+e_t
\end{aligned} \tag{C.31}$$

$$\begin{aligned}
y_t =&\beta X_t + n_t \\
n_t =&n_{t-12} + (n_{t-1}-n_{t-13}) \\
&+\theta_1 e_{t-1}+\theta_2 e_{t-2}+\Theta_1 e_{t-12}+\theta_1\Theta_1 e_{t-13}+\theta_2\Theta_1 e_{t-14}+e_t
\end{aligned} \tag{C.32}$$

Adding the estimated model coefficients from Table 4.16 gives:

$$\begin{aligned}
y_t =&2712 \times WkdyServMiles1000_t \\
&+5472 \times NumStations_t \\
&+0.2027 \times Emp4Cty_t \\
&+8098 \times SFEmpPct_t \\
&-20794 \times CashFare_t \\
&+86312 \times FuelCost_t \\
&-19010 \times StrikeDays_t \\
&+n_t \\
n_t =&n_{t-12} + (n_{t-1}-n_{t-13}) \\
&-0.5701e_{t-1}-0.2827e_{t-2}-0.6603e_{t-12} \\
&+0.3764e_{t-13}+0.1867e_{t-14}+e_t
\end{aligned} \tag{C.33}$$

Equation C.33 is the final equation for the application of the BART regression model with ARIMA errors.