2022

Accuracy and Uncertainty in Traffic and Transit Ridership Forecasts

Jawad Mahmud Hoque
jawadmhoque@uky.edu
Author ORCID Identifier:
https://orcid.org/0000-0002-8803-3899
Digital Object Identifier: https://doi.org/10.13023/etd.2022.155

Jawad Mahmud Hoque, Student

Dr. Gregory Erhardt, Major Professor

Dr. Lindsey Bryson, Director of Graduate Studies

Accuracy and Uncertainty in Traffic and Transit Ridership Forecasts

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Engineering
at the University of Kentucky

By

Jawad Mahmud Hoque

Lexington, Kentucky

Director: Dr. Greg Erhardt, Associate Professor of Civil Engineering

Lexington, Kentucky

2022

https://orcid.org/ 0000-0002-8803-3899

ABSTRACT OF DISSERTATION

Investments of public dollars on highway and transit infrastructure are influenced by the anticipated demands for highways and public transportations or traffic and transit ridership forecasts. The purpose of this study is to understand the accuracy of road traffic forecasts and transit ridership forecasts, to identify the factors that affect their accuracy, and to develop a method to estimate the uncertainty inherent in those forecasts. In addition, this research investigates the pre-pandemic decline in transit ridership across the US metro areas since 2012 and its influence on the accuracy of transit forecasts.

The sample of 1,291 road projects from the United States and Europe compiled for this research shows that measured traffic is on average 6% lower than forecast volumes, with a mean absolute deviation of 17% from the forecast. Higher volume roads, higher functional classes, shorter time spans, and the use of travel models all improved accuracy. Unemployment rates also affected accuracy—traffic would be 1% greater than forecast on average, rather than 6% lower, if we adjust for higher unemployment during the post-recession years (2008 to 2014). Forecast accuracy was not consistent over time: more recent forecasts were more accurate, and the mean deviation changed direction. Similarly for 164 large-scale transit projects, the observed ridership was about 24.6% lower than forecasts on average. The accuracy depends on the mode, length of the project, year the forecast was produced as well as socio-economic and demographic changes from the production to observation year.

In addition, we have found evidence of recent changes in transit demand to be affecting the transit ridership forecast accuracy. From 2012 to 2018, bus ridership decreased by almost 15% and rail ridership decreased by about 4% on average across the metropolitan areas in the United States. This decline is unexpected, because it coincided with the period of economic and demographic growth: indicators typically associated with rising transit ridership. We found that the advent of new mobility options in ride hailing services, bike and scooter shares as well as declining gas prices and increasing transit fares have the highest impact on ridership decline. Adjusting the ridership forecasts for these factors in a hypothetical scenario saw an improved transit ridership forecast performance.

Despite the advances in modeling techniques and the availability of rich travel data over the years, expecting perfect forecasts (where observations are equal to the forecasts), may not be prudent because of its forward-facing nature. Forecasts need to convey their inherent uncertainty so that planners and policymakers can take that into account when they are making any decision about a project. The existing methods to quantify the uncertainty rely on flawed assumptions regarding input variability and interaction and are significantly resource intensive. An alternate method is one that considers the uncertainty inherent in the travel demand models themselves based on empirical evidence. In this research, I have developed a tool to quantify the uncertainty in traffic and transit ridership forecasts through a retrospective evaluation of the forecast accuracy from the two largest available databases of traffic and transit ridership forecasts. The factors associated with the accuracy and the recent decline in transit ridership lead the formulation of quantile regression as a new method to quantify the uncertainty in forecasts. Together with a consideration of decision intervals or breakpoints where a project decision might change, such ranges can be used to quantify project risk and produce better forecasts.

KEYWORDS: travel demand forecast, forecast accuracy, uncertainty, transit ridership decline, quantile regression, reference class forecasting

Jawad Mahmud Hoque

(Name of Student)

03/08/2022

Date

Accuracy and Uncertainty in Traffic and Transit Ridership Forecasts

By

Jawad Mahmud Hoque

Dr. Gregory D. Erhardt

Director of Dissertation

Dr. Lindsey Sebastian Bryson

Director of Graduate Studies

May 10, 2022

Date

DEDICATION

To

My Family

ACKNOWLEDGMENTS

I thank the Almighty Allah for granting me the strength to stay on course for the last 5 years. I thank Him for granting me the love of my dear wife, Riana. I thank Him for everything.

Riana, where would I be without you? Thank you for bearing with my short tempers in the last 4 months. Thank you for your presence in the darkest hours, in the most trying times. Now, it is time for me to return the favor and you to finish up your dissertation.

My dear sister, Shemonty, kept me grounded with her innocent questions and weekly reports of her school. When you read this, know that I miss you very *very* much. I know you will excel at whatever you put your focus on. I am tremendously inspired by the work ethics of my father and my mother. Each day passes, and I appreciate their contributions to my well-being more than the day before. Raad, my dear brother, I know you are always with me. I love you. My new sister, Shupti, welcome to the family.

I am in deep gratitude to my in-laws: my mother, father, and sister-in-law. They are the quirkiest bunch I have ever met, and I love them for it. My dearest friends, Abir, Shuvo, Anit, Imon, Ratul, Romael, Munif, Amit, Minhaz, are the best at what they do, and they are my rock, my lighthouse and help I can count on any time of the day. I am tremendously lucky to be one of them.

I consider myself fortunate to have Dr. Greg Erhardt as my Ph.D. supervisor. He has been an inspiration and a source of strength throughout the last five years. As a professional and as an academic, not to mention as a human being, Dr. Erhardt is an ideal for me. He remains my advisor for life, and I know that I can always count on his help. I wish him all the best for his future endeavors, and I hope to work with him again.

I would like to express my sincerest gratitude to the members of PhD committee, Dr. Souleyrette, Dr. Chen and Dr. Bollinger. I have always heard that the oral qualifying and final doctoral exams are two things to dread, but in my experience, it has been a memory to cherish. They have been patient, accommodating and superbly helpful as I navigated through murky waters, unfamiliar concepts and getting generally lost and confused.

This research stands on great contributions from Dr. Erhardt, Dr. Watkins, Dr. Wachs and Mr. David Schmitt. I would especially like to acknowledge the contributions of Mr. Schmitt who has so kindly provided me the database he has been personally curating for several years now. I also want to thank Austin, Vedant, Eugene, Fahmida, Sneha, Shraddha, my fellow graduate students for their uncountable favors throughout the years.

I would also like to acknowledge the wonderful Bangladeshi student community here in Lexington. Everyone here has contributed, starting from carpooling to the groceries to fascinating discussions to relieve the pressure on Friday nights. I would especially like to acknowledge (in no particular order) Drs Asif, Tasnuba, Pavel, Prottasha, and soon be Drs Foteh, Swarna, Badhan, Nahiyan, Tonni, Tusher, Tibra, Selim, Pritam, and Musharrat. Live long and prosper.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

The primary goal of the research is to develop a tool to quantify the uncertainty in traffic and transit ridership forecasts. It achieves this goal through a retrospective evaluation of the forecast accuracy from the two largest available databases of traffic and transit ridership forecasts. This research also investigates the pre-pandemic decline in transit ridership across the US metro areas since 2012 to ascribe their effects on transit ridership forecast accuracy. The factors associated with forecast accuracy, road traffic and transit ridership, lead the formulation of quantile regression as a new method to quantify the uncertainty in forecasts.

1.1 Background

President Biden's Bipartisan Infrastructure Law allocates $550 billion in new federal investment in America's infrastructure: on roads, bridges, and major public transit projects as well as on zero-emission vehicles ("President Biden's Bipartisan Infrastructure Law" n.d.). Such large investments of public dollars will be informed by the anticipated demands for highways and public transportations or traffic and transit ridership forecasts. Forecasts form the basis of major decisions of a transportation project—approval of funds, environmental impact, choice of alternative, and the design of the pavement or transit system itself to name a few. Therefore, the inaccuracy in the forecasts can affect the greater economic recovery from the COVID-19 pandemic in addition to project success in terms of benefits to cost estimates. For several decades now, studies have investigated the extent of such inaccuracies with a view to improving the travel demand modeling practice.

Scholars have also identified reasons for forecast inaccuracy— errors in input data and inaccuracy in exogeneous forecasts (socio-economic and demographic projections, construction time), the use of overly simplistic forecasting methodology, and potential strategic misrepresentation stemming from political motivations (Bent Flyvbjerg 2005). Such revelations prompt the question: how can we make forecasts that are “good-enough”?

Despite the advances in modeling techniques and the availability of rich travel data over the years, expecting perfect forecasts (where observations are equal to the forecasts), may not be prudent because of its forward-facing nature. Forecasts need to convey their inherent uncertainty so that planners and policymakers can take that into account when they are making any decision about a project . Scholars have argued for moving away from the usual single-point estimates to a range of probable outcomes through constructing uncertainty windows around forecasts. Together with a consideration of decision intervals or breakpoints where a project decision might change, such ranges can be used to quantify project risk. If an actual outcome at the low or high end of the range would change the decision, that should be considered a warning flag. Further study may be warranted to better understand the risks involved, or decision makers may choose to instead select a project with lower risk. The purpose of this study is achieving this goal of creating good traffic and transit ridership forecasts to aid policy-planners into making informed decisions.

This section discusses literature pertaining to the three ideas that make up this study. I look at the role of forecasting in planning and discuss the importance of consistent and reliable forecasting for effective planning. This sets the context on the usefulness of forecasts and the forecast performance evaluation criteria proposed in past works. For the second part of this literature review I focus on reviewing existing research in assessing forecast accuracy to set the stage of my analysis. This section identifies the state of the art and the factors that influence forecast accuracy in the context of transportation planning. The last and most crucial part assimilates the literature and discusses how to get better traffic and transit ridership forecasts based on empirical evidence.

1.1.1 Role of Forecasting in Transportation Planning

Putting very simply, planning is the deliberative and disciplined approach of shaping the future. The American Planning Association (APA) defines planning as "[providing] a vision for the community today–and in the future". But this definition is all-encompassing: every action a person takes can be termed as planning. Mintzberg (1981) closes the boundary by defining planning as "programming"- in that it is not a tool to conceive an idea, rather an elaboration of the consequences of the intended strategy. Steiner (2010) finds planning to provide a linkage between the present and the future through the analysis of cause-and-effect consequences over time. The goal of planning, they say, is to bring about a desired future through informed decisions. We can identify several important characteristics of planning: it starts with setting goals and objectives, identifying and analyzing the alternatives and articulating their consequences, setting the bounds and expectations, and finally designing the approach to achieve the goal.

In the public sector, planning occupies an integral and ongoing part of policymaking. Planners today confront issues concerning a myriad of geographic and socio-economic development like land use, public health, economic development, environmental risk mitigation etc. Plans communicate to the stakeholders— policymakers and the general public alike— a vision for addressing the present problem and how to achieve it (Meyer and Miller 2001). But often these problems lack a clear solution, and because of their forthcoming nature, the eventual solutions themselves are subject to much speculation and assumption on the part of the planners. Moreover, funding constraints particularly for massive infrastructure projects means that project decisions need to be reinforced by "*proper, systematic and neutral information*" (M Wachs 1985). Making a decision, for this reason, hinges on the analysis of anticipated benefits from several alternatives against the cost to establish the best course of action. In transportation planning, such alternative analysis through demand forecasts is required by law: the Federal Aid Highway Act of 1964 requires the highway plans to be evaluated against 20-year travel demand forecast (Martin Wachs 1990). Similarly, the Federal Transit Agency requires the state and federal agencies to submit alternate analyses for their proposed projects to be eligible for federal funding through the Capital Investment Grants Program, otherwise known as the New Starts Program.

The Fixing American's Surface Transportation Act ("FAST Act"), signed by President Obama in December 2015, provides $41.5 billion each year in roadway and bridge funding (U.S. Department of Transportation, Federal Highway Administration n.d.). How these public dollars are invested, depends on the anticipated demands for highways and public transportations through environmental impact assessments, benefit-cost analyses, capital cost estimates etc. The forecasts also directly influence the design of the facility: number of lanes on a proposed roadway, service frequency for a public transit route, estimate revenue for a toll road etc. Inaccuracy in forecasts therefore skew the benefit-cost estimates and may result in the selection of an alternative with less benefits or even inadequate design. This is particularly true for toll-road forecasts where the inaccuracy has a greater bearing on project success. As an evidence to this, the Australia

Government (2012) cited ‘‘*inaccurate and over-optimistic*’’ traffic forecasts as a threat to investor confidence. As (Bain 2009) put it, “*aggressive financial structuring leaves little room for traffic usage to depart from expectations before projects experience distress and debt repayment obligations become threatened*”. Three lawsuits are now underway that challenge the forecasts for toll road traffic which subsequently came in significantly under projections (Bain 2013). The consulting firms that produced the forecasts have settled these lawsuits with upwards of 80 million Australian Dollars (“Arup Settles $1.7B Australia Toll Road Revenue Forecast Suit” n.d.).

It is therefore quite apparent that good forecasts lead to good decisions. But what is a good forecast? Is it simply the forecast that predicts future outcome at pin-point accuracy? To spin the question differently, how can we evaluate the goodness of forecasts?

1.1.2 Establishing the criteria of good forecasts

The importance of forecasts in planning warrants that a perfect forecast would be the one where the actual observation exactly matches the forecast. But since the future is a moving target, such standards are often quite impossible to achieve, particularly in social science, economics and finance. This imperfection in forecasts is prevalent in natural sciences like climatology as well where the factors affecting change are numerous and sometimes unaccounted for in theory. The accuracy of forecasts thus cannot be the sole metric for evaluating the goodness of forecasts.

In *Forecasting, An Appraisal for Policy Makers and Planners* (Ascher 1979) Ascher evaluates forecasts made in different fields like energy, population, economic, transportation etc. and identifies the “insider’s approach” and the “outsider’s approach” as different point of views to evaluate forecasts. Insider's approach, in his words, focuses on the appraisal of the forecasting technique and outsider's approach focuses on the accuracy of the forecasts as a whole. Forecasts for strategic planning are very much dependent on external factors and the accuracy of the forecasts is influenced by the forecast's ability to absorb the uncertainty in the external factors. While improving the forecasting technique itself is necessary, it should be balanced by the proper use of forecasts based on limited information about the external factors (Naylor 1983). According to (Naylor 1983), the balance of the two approaches described by Ascher comes in the form of multi-scenario or "What-if?" forecasts and sensitivity analysis to ascertain the effect of deviation from the forecast. So, the first characteristic of a good forecast is a representation of the uncertainty.

Naylor's proposition is expanded upon in subsequent research particularly in the field of weather forecasting. (Murphy 1993) identifies that the forecast performance can be looked at from two different perspectives. For the forecaster themselves, this evaluation means how much the observed condition matches the forecast condition. For the user of the forecasts, a good forecast simply refers to their utility in getting a beneficial outcome. The author proposes three metrics for evaluating the goodness of forecast: consistency, quality and value. According to this study, forecasts need to correspond to the knowledge base of the forecaster (consistency of forecasts) by expressing the uncertainty inherent in the forecasting process through probabilistic terms. The joint distribution of forecasts and observation expresses the time-invariant information relevant to the correspondence of the forecasts to the observations (quality of forecasts). The benefits of forecasts to the users (value of forecasts) is determined by the decision-making characteristics, e.g. course of action, payoff and the quality of the forecast itself etc.

Regarding the value or usefulness of forecast, (Murphy 1993) states that the forecasts themselves don't have any intrinsic value, rather they derive their value from their use in decision-making. (Voulgaris 2019b) proposes two qualitative measures to establish the value or usefulness of forecasts: whether the decision would change for a different forecast and would the unselected decision lead to a better outcome.

(Voulgaris 2019b), taking que from forecast studies in various disciplines like politics, meteorology, public health and transportation planning developed a forecast evaluation framework. According to the author, the three characteristics that are relevant for forecast evaluation are methodology, accuracy and usefulness. Evaluating the forecast methodology (i.e. inputs, assumptions, mathematical models) is similar to the insider's approach as advocated by (Ascher 1979). The accuracy and bias of point forecast is relatively simple to evaluate, although the interpretation is dependent on which metric is used. The author describes several measures like Mean Absolute Percent Error (MAPE) and Mean Magnitude of Error Relative to the Forecast (MMER) as useful measures for assessing the accuracy of point forecasts. Such measures correspond to Ascher's "outsider's approach" as they measure the accuracy of the output, rather than the inputs and methodology. But, as (Mason and Stephenson 2008) points out, such single metric fails to convey all the important information about the forecast quality because of its multi-faceted nature. The authors instead suggest using relative frequencies of forecast and observation to evaluate the quality of forecasts. These measures can be absolute (reliability), categorical (resolution) and conditional (discrimination) and can be used in conjunction with accuracy.

For demand forecasts to be "good", this framework provides a useful guideline for getting good forecasts:

- The methodology should incorporate accurate inputs and better mathematical models in addition to correct assumptions regarding future condition.
- The uncertainty inherent in the forecasting process needs to be conveyed by the forecast through probabilistic terms.
- The relative frequencies of actual and forecasted demand (traffic and transit ridership) will establish the quality of the forecasts.
- Forecasts need to aid decisions in such a way that extremities in the range do not change the decision. In other words, policy decision should consider the maxima and minima of a range of forecasts thereby establishing the value.

The purpose of the current study is to get good travel demand forecasts based on the evaluation criteria described above. Traffic forecasting is a model of short or long term aggregated human behavior in the presence of a stimuli like a newly developed mode of transport, an expanded roadway or a new bus route in an existing network. Transportation planning agencies estimate demand for these and other scenarios and alternatives. But the elasticity of such estimates or forecasts with respect to the inputs to the model makes accuracy a difficult goal to attain. It is very challenging to anticipate, or even identify, all the factors that can potentially affect travel behavior. This uncertainty is further convoluted by the presence of external factors beyond the control of the planner– the political economy in particular (Brooks 2019; Martin Wachs 1990). Demand forecasts in this sector are, however, mostly point-estimates rather than probabilistic range, and are typically subject to significant variability from observation. To gauge the feasibility of generating better forecasts by addressing the uncertainty, we need to take a look at the previous studies in travel demand forecast accuracy.

1.1.3 Past Research in the Assessment of Forecast Accuracy

Accuracy of traffic forecasts have been a point of concern for several decades now. Limited availability of funds for transportation projects coupled with the potential impact of inaccuracy have enabled this inquiry to garner more attention in recent years. Even so, the number of probes into this topic have been few and far between.

Investigations by Melvin M. Webber on San Francisco's construction of the Bay Area Rapid Transit (BART) system was one of the first examples of an in-depth analysis of traffic forecast (Webber 1976). Webber compared the actual daily usage of the system as well as the effect on auto-ridership to the predicted. Webber's analysis found significant deviation of the actual scenario from the forecast. The total patronage of the system (average weekday trips) in 1976 was about half of what was predicted for 1975.

Similar to the analysis on BART, Professor Kain looked into the Dallas Area Rapid Transit (DART) in 1990 (Kain 1990). He found that DART made extensive use of unrealistic land use forecasts and optimistic ridership forecasts to obtain voter approval for a 91-mile rail transit system. Although not exactly an examination into the accuracy of the forecasts, the author instead focused on the appraisal of the techniques employed (i.e., an example of the insider's approach proposed by Ascher). According to the author, the most serious error in developing the long-term transit plan was the lack of alternative analyses, as well as using flawed land-used projection and highly optimistic ridership forecasts.

The number of forecasting accuracy assessments have increased since the year 2000, with several focused on assessing the accuracy of toll-road forecasts. The inspiration seems to be from the fact that toll road forecasts have a bearing on investor expectations and that is why their accuracy is more important.

The general inaccuracy and optimism bias in demand forecasts were observed first in 2005 (Bent Flyvbjerg, Skamris Holm, and Buhl 2005). The conclusions were based on over 210 transportation projects (27 rail projects, 183 road) from across the world. The authors found that rail ridership forecasts are less accurate and more inflated than road vehicle forecasts at a very high level of statistical significance. The researchers found at least 25% of the road projects go beyond the ±40% error range and about 50% stray beyond ±20%. The researchers also could not identify any evidence to the claim of increasing accuracy over time through statistical tests. The study identified inaccurate assumptions and exogenous forecasts (tied to the concept of optimism bias), deliberately slanted forecasts, issues with the analytical tools and issues with construction or operation as contributing factors. As a follow up to this study, (Bent Flyvbjerg 2013) discusses the systematic misrepresentation of forecasts, underestimating costs and overestimating benefits, in large infrastructure projects. The author advocates taking an “outside view”: using experience from similar ventures already completed for getting reasonable estimates not biased by inside information and any other socio-political incentives.

Assessments of toll-road forecasts (Bain and Polakovic 2005; Li and Hensher 2010) also confirm the evidence of optimism bias. (Bain and Polakovic, 2005) analyzed 104 toll road projects and found this bias persisting into initial five years of operation. The factors the researchers identified as drivers behind this bias were mostly the toll culture (existence of toll roads previously, toll acceptance etc.) and errors in data collection as well as unforeseen micro-economic growth in the locality. These findings went on to become the basis of Standard & Poor's Traffic Risk Index, an empirically derived risk register for investors and financial analysis (Bain, 2009). Similar observation on optimism bias in toll road forecasts in Australia is reported in (Li and Hensher 2010). The researchers found actual traffic for the roads were about 45% lower than the predicted value on an average in the first year of operation. The accuracy doesn't get better over time, as the percentage error reduces by only 2.44% each year after opening. They attributed this error in forecast to less toll road capacity (when opened, compared with forecast), elapsed time of operation (roads opened longer had higher traffic levels), time of construction (longer construction time delayed traffic growth and increased the error), toll road length (shorter roads attracted less traffic), cash payment (modern no-cash payment increased traffic), and fixed/ distance-based tolling (fixed tolls reduced traffic). At the opposite end of the spectrum, inaccuracy-wise, lies toll road forecasts in Norway. Odeck and Welde (2017) investigated 68 Norwegian toll roads and found that while toll-road traffic is underestimated, they are close to accurate as the mean percentage error is a mere 4%. They attributed the standard organizational framework of a national toll forecasting system with "*little or no incentives to exaggerate the forecast*" as a factor.

Similar to the accuracy of toll road traffic forecasts, transit ridership forecasts have also attracted attention over the years. The BART and DART analyses (Webber 1976; Kain 1990) are examples of researches into this aspect. In more recent times, the Federal Transit Administration (FTA) has conducted several studies analyzing the predicted and actual outcomes of large-scale federally funded transit projects (Lewis-Workman et al. 2003; 2007; Federal Transit Administration 2020). The FTA is finding that transit forecasts are becoming more accurate over time, and attribute that improvement to better scrutiny of travel forecasts and the analytical tools used to produce the forecasts.

Schmitt (2016) presented the results of his analysis of 65 large-scale transit infrastructure projects New Starts built in the United States through 2011. The research found that transit project assumptions have historical bias towards over-forecasting ridership. (Voulgaris 2019a) analyses the accuracy of transit ridership forecasts of the 67 projects in the same database used by Schmitt against several explanatory variables like project characteristics, time between forecast and observation, local experience with project mode and physical and financial characteristics. The author found that transit forecasts, on average, are biased but have been becoming more accurate over time. The strong correlation between forecast accuracy and project mode was also observed.

Compared to the analysis of accuracy for toll roads and transit projects, studies into non-tolled roadways are few. Most of the studies have been limited in scope as well, assessing the performance of state-wide models or MPO's forecasting tools. (Anderson, Vodrazka, and Souleyrette 1998) evaluated the performance of Iowa travel demand model for two projects. The research revealed that poor estimates of horizon year demographic and socio-economic data contributed most to the errors in the forecast. Parthasarathi and

Levinson (2010) examined the accuracy of traffic forecasts for one city in Minnesota and found the mean error to be 8%. In this study the researchers took the mean of the error values which can be positive or negative. Since positive and negative errors offset each other, this statistic only gives the mean of the distribution, rather than any absolute measure of the deviation of the actual traffic. Giaimo and Byram (2013) examined the accuracy of over 2,000 traffic forecasts in Ohio produced between 2000-2012. They found the traffic forecasts slightly over-predicting, but within the standard error of the traffic count data. (Buck and Sillence 2014) evaluated 131 forecasts in Wisconsin and determined the mean absolute difference between the forecasted and actual traffic to be 16%. In the study of 39 road projects in Virginia, Miller et al. (2016) reported that the median absolute percent error of all studies was about 40%. The percent error values in this study is higher than those reported in (Buck and Sillence 2014; Flyvbjerg, Holm, and Buhl 2005). This study also quantifies how certain factors affect the forecast accuracy. According to their research such factors are- Forecast Method (trend based more accurate than activity based under a few conditions) and forecast duration (as it decreases, accuracy increases).

Nicolaisen (2012) measured the forecast inaccuracy for 146 road projects in Denmark, Norway, Sweden and the UK and found that around two-thirds of the projects have observed traffic volumes that fall within ±20% of the forecast. Forecasts were biased towards under-prediction. Limitation in the data made investigating the indicators of forecasting accuracy difficult. But the author found no clear evidence of improvement in forecast accuracy over time for road projects. He also found less errors in forecasts for upgrading existing roadways than that for new links. The author hypothesizes that poor traffic distribution models may be more at fault than overestimation of actual traffic demand for the inaccurate forecasts.

It is quite apparent from past studies that travel demand forecasts are generally inaccurate and optimistically biased. Yet, the inherent uncertainty in the process are not addressed and forecasts are still provided as point-estimates. (Hartgen 2013) called this a representation of the forecaster's "hubris" and called for range-based forecasts to be the industry standard. (Bain 2011) too proposed creating "uncertainty envelopes" around forecasts to address this issue. The question then becomes, how to do it.

1.1.4 The Conceptual relationship between Forecast Accuracy and Uncertainty

Most of the existing research on forecast accuracy prescribe getting better data on the inputs, modifying assumptions to incorporate more recent and complete travel behavior in the demand model in addition to advancing the model itself. (Ascher 1979) demonstrated that better input data produces better forecasts, even more so than better and more complex models. But if the uncertainties around these input values are not addressed, they inevitably propagate through even the most complex travel demand models and result in forecast error (Zhao and Kockelman 2002). Synthesizing the literature, we can identify two problems with input data to demand models:

1. They are cross-sectional and thus only represent a static travel behavior.
2. The variance in exogenous forecasts to demand models (socio-economic and demographic forecasts) are not incorporated in the models, rather only the point estimates are used.

Simulating the uncertainty in inputs and conducting sensitivity tests by running the model with extremities in inputs can address some of the uncertainty. But they inevitably come with the cost of upgrading and maintaining the modeling process in addition to added runtime. In addition, such sensitivity testing are not useful for forecasts done using a simple trend-line analysis. (Voulgaris 2019b) suggests that forecasting resources are better spent on simple models and averaging or combining results.

The other option of producing better forecasts is employing what (Ascher 1979) calls as “outsider’s approach” and Kahneman and Tversky (1977) calls “reference class forecasts”. Reference Class Forecasting is the use of the base-rate and distribution results from similar situations in the past to improve forecast accuracy. The benefits of reference class forecasting were suggested in subsequent work by Flyvbjerg (2007) and Schmitt (2016) to correct for biases in demand and cost forecasts. Flyvbjerg suggested developing and applying reference classes to projects with large uncertainties to get more accurate forecasts. (Bain 2011) interviewed industry professionals (consultants, modelers, academics) to gauge their estimates of uncertainty in their forecast and prepared an “uncertainty envelope” (Figure 1) that can provide a baseline for converting point forecasts to range-based ones.

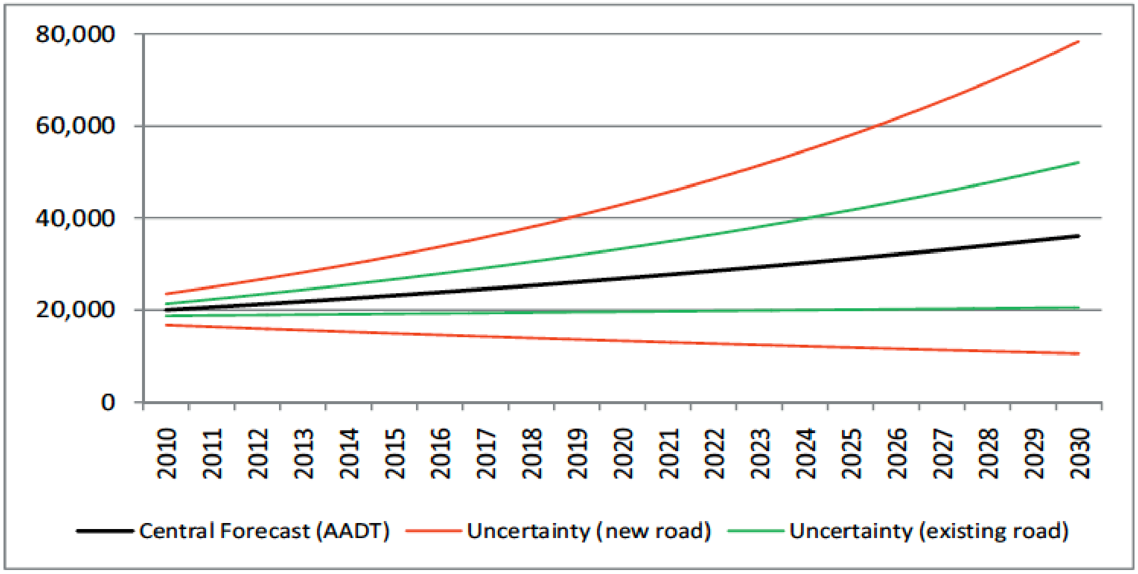

Figure 1: Uncertainty Envelope, Source: (Bain 2011)

Using the principle of Reference Class Forecasting, such uncertainty envelopes around forecasts can be constructed from empirical evidence: accuracy assessment of a large enough sample of forecasts and observations of traffic and ridership. Separate reference classes can be established for distinct categories provided large enough sample sizes, for example- traffic forecasts by roadway functional class or project type (new construction, existing roadway) and transit ridership forecasts by locality type (transit or auto oriented, high, or low population density) or project type (rail or bus route development) etc. Uncertainty envelope created in this manner takes care of the reliability criteria as well, since the range of values would represent the percent of observations that fall within. The assessment of forecast value will be then determined by the possibility of changing a project decision if the actual observation is at the high or low end of the envelope.

1.2 Research Objective

As demonstrated in the previous section, despite understanding the necessity of expressing traffic and transit ridership forecasts as probabilistic ranges rather than point-based ones, there haven't been extensive studies in this domain. The hurdle seems to be the lack of appropriate data to conduct statistically significant analysis into the extent of inaccuracy and factors affecting such inaccuracy (Nicolaisen and Driscoll 2014). Recently at least two databases with large sample size have been created for storing project level forecast and actual observation information (traffic forecasts in Erhardt et al., 2019; transit ridership forecasts in Schmitt, 2016) allowing rigorous statistical analysis. Using these databases, a possibility has opened up to quantify the uncertainty in forecasts.

There are multiple mechanisms for quantifying uncertainty in demand forecasts. One method to quantify uncertainty is sensitivity testing: varying the forecast inputs and assumptions to reflect their uncertainty ranges and re-run the travel demand model with multiple inputs. This process can be repeated many times, so that all primary inputs vary by their (minimum and maximum) extreme values individually and collectively. The result is a distribution of outcomes reflecting the specified range of inputs and assumptions. This method is less pragmatic if the travel model has long running times, project schedules are constrained, or if a simple a trend line extrapolation was used to produce the point-forecast. In addition, running the same model multiple times with different input assumption do not necessarily get rid of the technical limitations of the model. These situations commonly occur in travel demand forecasting, so an alternative method is needed.

To achieve the goal of generating good and useful forecasts, I will be presenting in this study a novel means of estimating the range of uncertainty around a forecast using a technique called quantile regression. The quantile regression models are estimated from the actual demand as a function of the forecast and provide a means of predicting the range of expected demand from a single forecast. The ranges produced by this method are empirical, meaning that they consider the full set of possible errors that have occurred in the past, rather than leaving it up to the analyst to determine a reasonable range of inputs. This is both an advantage and a disadvantage. It is beneficial because it may implicitly incorporate factors that the analyst may not consider on their own. However, it is limiting if the future looks very different from the past. For example, a risk in forecasts made in 2019 may be the effect of self-driving vehicles, and that risk is not one that has been an issue for projects that are already open. Another advantage is that employing such models would require a lot less time and computing power. To obtain a probably range of outcomes for a particular point forecast, it is as simple as tracing lines on a chart or inputting values to a spreadsheet. The models themselves can act as a performance metric for an agency since they incorporate observations and predictions in a unified model.

For the models to be useful for agencies, they need to incorporate the recent changes in travel behavior, not limited to the COVID-19 pandemic itself. Studies have observed that across the United States, public transit ridership has declined unexpectedly from 2012 to 2018. Factors that have contributed to this decline may have some effect on ridership forecast as well. In the third part of this study, I will be exploring the factors associated with the decline in transit ridership to investigate their effect on forecast uncertainty. The more specific objectives of this study are:

- To establish empirical evidence of uncertainty in traffic and transit ridership forecasts
- To identify factors affecting the uncertainty in traffic and transit ridership forecasts and
- To develop quantile regression models to quantify the uncertainty in these forecasts.

1.3 Thesis Structure

As previous studies establish, the impediment to conducting statistically significant analysis into post-opening forecast evaluation is the lack of available data. In chapter 2 of this dissertation, I briefly describe the two databases assembled to quantify the accuracy of traffic and transit ridership forecasts. In the latter part of the chapter, I present the theoretical relationship between accuracy and uncertainty and discuss the research approach pertaining to assessing the accuracy of traffic and transit ridership forecasts and the quantile regression methodology.

The accuracy of traffic forecasts is presented in Chapter 3. It discusses the general state of accuracy in traffic forecast: overall and by several descriptive categorical variables. The variables selected are based on available fields in the database and prior research in this area. The chapter also discusses the effects of the Great Recession from 2008 to 2012 on the accuracy of road traffic forecasts for projects that opened during this period. The chapter has been adapted from (Hoque, Erhardt, Schmitt, Chen, Chaudhary, et al. 2021b).

In chapter 4, I go over the recent decline in public transit ridership in more detail. The analysis is based on a dataset of 215 metropolitan statistical areas (MSAs): their corresponding fixed-route transit ridership, socio-economic and demographic characteristics, and data on several explanatory variables. The variables chosen are based on an extensive literature review of factors affecting public transit ridership and I present a brief review of it in this chapter. The analysis results quantify the effect of each variable in the final econometric model on transit ridership decline between 2012 and 2018.

Chapter 5 follows a similar approach to chapter 3 but discusses the state of accuracy in transit ridership forecasts instead. Here, I reference the declining transit ridership trend from 2012 to 2018 and discuss their potential effect on the accuracy of the sample.

I present the results of the quantile regression models for both traffic and transit ridership forecasts in chapter 6. The chapter also demonstrates the application of the regression models.

The thesis finishes with conclusions, lessons learned and the next steps. These relate to the overarching goal of improving the forecasting practice by acknowledging the uncertainty inherent.

Chapter 2 RESEARCH APPROACH

The goal of this study is creating a set of tools for forecasters to generate better traffic and transit ridership forecasts. As described in the previous section, analysis of previous studies in this domain allows to develop more specific objectives: establishing the empirical evidence of uncertainty in the forecasts and creating uncertainty envelopes using the evidence. This chapter details the research approach employed to achieve these objectives. I discuss the databases created for the three distinct parts of the research: traffic and transit ridership forecast accuracy and factors affecting transit ridership decline. The chapter also presents the theoretical background of quantile regression as a tool for quantifying uncertainty. The general methodology for quantifying forecast accuracy and the effect of different explanatory variables on transit ridership decline are discussed in brief here.

2.1 Introduction

The literature review identifies that most of the works in the domain of forecast accuracy assessment was either a statistical exploration of a large sample of data, or an in-depth analysis of the forecast performance of a single project or forecasting model. Both have their advantages and disadvantages— the first approach allows us to evaluate the method itself as a whole and the second allows us to appraise a particular model and go into minute details about its weaknesses. For this study, where the goal is understanding and quantifying the uncertainty in travel demand forecasts, the statistical analysis of a large sample is appropriate. This approach relies on gathering a large sample of forecasts for which data were collected and the forecasts were made sufficiently long ago that the horizon year of the forecasts has come. This makes it possible to compare the forecasts of demand with measured demand on the facilities for which the forecasts were made. With a large sample of such forecasts, we use statistical analysis to examine correlations between forecast accuracy and data inputs, facility types, methods used to conduct the forecasts, and factors exogenous to the forecasts that influenced their accuracy.

However, in his review of the 50-year history of travel forecasting, David Hartgen (2013) said, "*The greatest knowledge gap in US travel demand modeling is the unknown accuracy of US urban road traffic forecasts.*" Researchers have improved travel demand forecasting methods in recent decades but invested relatively little in understanding their accuracy. This underinvestment is unfortunate because accurate forecasts improve decisions about the evaluation, selection, and design of transportation projects. The absence of data has been the major barrier to the study of travel forecast accuracy (Nicolaisen and Driscoll 2014). This deficiency arose because accumulating the data needed for retrospective analysis requires proactive planning. The responsible agencies do not commonly preserve and archive forecasts, and so often lose these data. Long project development cycles and staff attrition make recovering this information cumbersome.

The second challenge in quantifying the uncertainty is reconciling it with accuracy assessment. Evaluating accuracy is a retrospective activity that accounts for past forecast errors, while expressing uncertainty is a prospective activity that considers possible errors. In this chapter, I articulate the relationship between accuracy and uncertainty, and propose these empirical measures of past forecast accuracy as an estimator of the uncertainty in future forecasts. The method, Quantile Regression, starts from an econometric framework that uses Ordinary Least Squares regression to model measured demand as a function of forecast demand for the purpose of detecting bias. I extend that framework in two ways: first, to measure the spread of outcomes in addition to the bias, and second, to measure the effects of exogenous predictors on both bias and spread. This method differs from traditional methods of estimating uncertainty which rely on assumptions about reasonable ranges of travel demand model input values and parameters. This represents a significant advance in the methods for the study of forecast accuracy. It is useful to researchers who wish to understand the variables associated with forecast accuracy, such as whether some forecasting methods are more accurate than others.

2.2 Data

For quantifying the uncertainty in traffic forecasts, we employed a database created as part of the National Cooperative Highway Research Program (NCHRP) Project titled Traffic Forecast Accuracy Assessment Research (G. D. Erhardt et al. 2019). The database currently contains forecast information from the six participating states (Florida, Michigan, Minnesota, Ohio, Massachusetts and Wisconsin) as well as four European countries (Denmark, Sweden, Norway and the United Kingdom, obtained from Nicolaisen (2012)). The sources are the DOT maintained databases, ESAL reports, project forecast reports and/or traffic/environmental impact statements as well as database from similar research efforts. The database contains information on the project itself (unique project ID, improvement type, facility type, location), forecast (year forecast produced, forecast year, methodology etc.) and the actual traffic count information.

The second database to be used for evaluating uncertainty in transit ridership forecasts is collected courtesy of Dave Schmitt, first demonstrated in (D. Schmitt 2016). This database currently contains information on 142 transit projects funded by the New Starts program by FTA. The Transit Forecasting Accuracy Database (TFAD) contains detailed information on the demand forecast accuracy of large-scale transit projects in the United States. Characteristics of the projects are also included, such as length, number of stations, and whether the project services the region's Central Business District.

2.2.1 Traffic Forecast Database

We compiled a database containing forecast information from six participating states (Florida, Michigan, Minnesota, Ohio, Massachusetts and Wisconsin) and four European countries (Denmark, Sweden, Norway and the United Kingdom). The sources included Department of Transportation (DOT) databases, project forecast reports and/or traffic/environmental impact statements as well as databases from other published studies (Parthasarathi and Levinson 2010; Nicolaisen 2012; Giaimo and Byram 2013; K. Buck and Sillence 2014b; Marlin Engineering 2015; Miller et al. 2016). The database includes information on each project (unique project ID, improvement type, facility type, location), forecast (forecast horizon year, methodology etc.) and the post-opening traffic count information. The data contain a diversity of projects, including new roads, road widenings, interchange reconstructions, safety and operational improvements and pavement resurfacings.

In total, the database contains reports for 2,611 unique projects, and for 16,697 road segments that comprise those projects. Some of the projects had not yet opened; some of the segments did not have traffic count data associated with them, and others did not pass the quality control checks for inclusion in statistical analysis. While we retained all records for future use, we based our analysis on a subset of 1,291 projects and 3,912 segments, as Table 1 shows. Most participating agencies compiled the data retrospectively. For some jurisdictions, we only have data for a few projects, and those projects tend to be larger in scope and therefore better documented. In contrast, for Agency E, we have data for nearly every forecast made starting in the early 2000s, comprising 44% of our sample. (Because their management directed them to clean their office, staff from this agency also provided about two dozen boxes of paper records for these projects. We organized and digitized the basic attributes of these projects and included them in our database.)

Table 1: Traffic Forecast Database Summary of Available Data

Jurisdiction	All Projects		Open Projects with Required Data	
	Number of Segments	Number of Unique Projects	Number of Segments	Number of Unique Projects
Agency A	1,123	385	425	381
Agency B	12	1	12	1
Agency C	38	7	6	3
Agency D	2,176	103	1,292	99
Agency E	12,413	1,863	1,242	562
Agency F	463	132	463	132
Agency G	225	73	225	61
Agency H	23	23	23	13
Agency I	21	10	21	10
Agency J	203	36	203	29
Total Segments	16,697	2,611	3,912	1,291

As Flyvbjerg (2005) recommends, we evaluated opening-year conditions. We defined the opening-year as the first post-opening year with traffic count data available. If we had multiple forecasts for a single project (such as opening-year and design-year forecasts, usually 20 years after project opening), we used the forecast closest to the opening-year. To make the comparison in the same year, we held the counts constant and scaled the forecast to the year of the count using the growth rate implied by opening and design year forecasts, and a standard growth rate of 1.5% if they were unavailable.

Projects did not always open in the year anticipated. This happened if a project was delayed, if a forecast was for an alternative design that was not built, or if funding priorities changed. We usually knew when delays occurred for large projects. For smaller projects we could not always determine when and if construction finished because DOTs do not necessarily link forecast records to construction records. Where we could not verify the project completion date (for 488 projects out of the 1291 in our analysis), we assumed that maintenance, minor construction or low risk projects were completed within one year of planned opening, and that major construction projects took two years beyond that. This assumption reduced the risk of including counts collected prior to the project opening. Because most projects took place on existing facilities, pre-opening counts are often available but may be affected by construction activity.

When comparing forecasts and counts, we compared the Average Daily Traffic (ADT), although its exact definition depended upon the source. Some agencies provided data as Average Annual Daily Traffic (AADT), some as Average Weekday Daily Traffic (AWDT), and some as typical weekday traffic, which usually was for non-holiday weeks with school in session. The units were not always clear in the data, so they may vary between agencies, but we assumed consistency between forecasts and counts within an agency.

In addition to the forecast and counted traffic volumes, we compiled the attributes for each project as Table 2 shows. Not all attributes were available for every project, usually because those data were not recorded when the forecast was made.

Table 2 also indicates the percent of projects with each attribute available. Different agencies also have different practices for recording attributes such as Functional Class, Improvement Type and Forecast Method, so we mapped those attributes to common categories.

We compiled these data based on their availability and they do not represent a random sample of transportation projects. We analyzed projects opening between 1970 and 2017, with about 90% opening in 2003 or later. We do not have details about the nature and scale of some projects, but earlier projects were often major infrastructure capital investment projects and later projects were often routine resurfacing projects on existing roadways. This trait of the database occurs because some state agencies began routine tracking of all forecasts only within the past 10 to 15 years and, in earlier years, retained only information for major investments. Similarly, the type of project, the methods used, and the specific data recorded all differ because of the practices of the agencies providing the data.

Table 2: Traffic Forecast Database Data Fields and Definition

Variable Name	**Description**	**Percent Available**
Forecast	Forecast daily traffic.	100%
Count	Counted daily traffic.	100%
Agency Type	Whether the forecasting agency is a State DOT, MPO or consultant	56%
Agency	Geographic location of project by State/Country	100%

Variable Name	Description	Percent Available
Functional Class	FHWA specified functional classification of the roadway	72%
Area Type	The area type where the facility lies: Rural, Mostly Rural, Urban and Unknown area types according to US Census Bureau's definition of Urban and Rural areas. The Bureau defines urban areas as a territory that has at least 2,500 people. The percentage of people living in rural areas in a county determines whether the county is rural (100%), mostly rural (50-99%) or urban (<50%).	91%
County Population Growth	Percent change in population between start year and forecast year. Stable counties are defined as having growth rate between -1% and 1%, declining counties have greater than 1% decrease and growing counties have greater than 1% increase in population.	73%
Improvement Type	Type of project: improvement on an existing roadway, new construction project.	72%
Forecast Method	Methodology for forecasting: using travel demand model, population growth rate, traffic count trend, professional judgement.	48%
Start Year	The year when forecast was produced.	100%
Forecast Year	The forecast horizon year. Sometimes our data include both opening-year and design-year forecasts for the same project, but we limit our analysis to opening-year conditions.	100%
Opening Year	The earliest year after project opening that traffic count data are available.	100%
Forecast Horizon or Time Span	Number of years between start year and opening year.	100%
Unemployment Rate in the Start Year	County level unemployment rate in the start year, obtained from the Bureau of Labor Statistics. For European projects, the national unemployment rate was obtained from the World Bank historical unemployment rate data	100%
Unemployment Rate in the Opening Year	County level unemployment rate in the project opening year.	100%

2.2.2 Transit Ridership Forecast Database

According to Transport Politic, approximately 283 unique projects have been constructed between 1974-2019 in the United States ("The Transport Politic - Transit Explorer 2021" n.d.). We based our analysis of transit ridership accuracy and uncertainty on a database of 164 large-scale transit projects across the United States. The database is compiled through personal efforts by Mr. Dave Schmitt (D. Schmitt 2016) and is currently the largest known database of this kind. The projects include downtown people movers, Bus Rapid Transit (BRT), Light Rail Transit (LRT), Heavy Rail Transit, and commuter rail. Information contained in the database include, but are not limited to, project and forecast characteristics like length, location, mode, service area and travel time characteristics, observed ridership where available and exogeneous forecasts like cost estimates, population, and employment projections etc. In addition, we have also made use of the set of projects included in (Voulgaris 2019a) to fill out missing fields and add more projects in the dataset. Table 3 presents the summary of data fields used in this study.

As we know from the project development life cycle, forecasts are made at different phases in the planning process. In our database, we have several ridership forecasts made at different project development phase. For consistency, we considered the forecast at the latest available stage of the cycle. Most often, this is the funding decision phase, as the forecast for the design phase are typically optimistically biased to avoid under-designing. For apples-to-apples comparison, the forecast and observed ridership needs to be in the same year as well. In case the observation is at a later year, we interpolated the forecast using to be at the same year as observation. After applying such selection criteria, we based our analysis on a reduced sample of 125 projects, all of which has an observation and a forecast ridership in the same year.

A limitation of the database is the high degree of missing data on key variables. Because of the absence of standardized reporting of project and forecast information, such data are often not recorded in the project documents released to the public. The projects span five decades, from the 1970s to the 2010s. Projects built since 2000 comprise over 70% of the database (Figure 2).

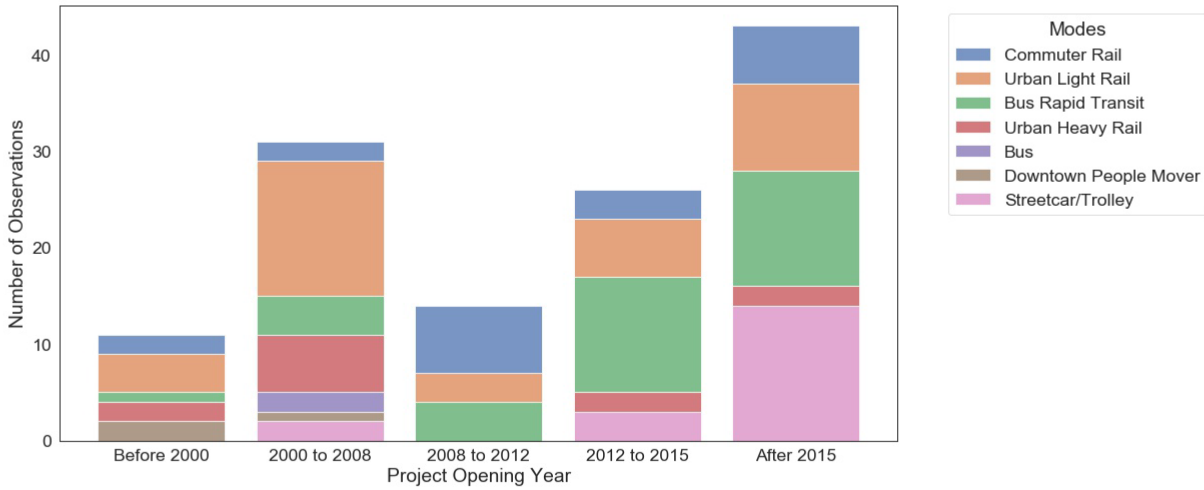

Figure 2: Number of Observations in the Database by Mode and Project Opening Year

The socio-demographic data have been collected at the Metropolitan Statistical Area (MSA) level from the American Community Survey (ACS) data, and the Bureau of Labor Statistics (BLS) data. However, the MSA delineation have changed over the years and data before 2005 and after 2019 were not available at the time of analysis. In such case, we used linear interpolation from the decennial census data to fill the blank fields. Such interpolation introduces additional bias in the analysis as these are different from the data used in the models. However, they do present the opportunity to evaluate the changing accuracy as the demographics shift over the years.

Table 3 Transit Ridership Forecast Database Data Fields and Definition

Field	Definition
Forecast Ridership	Forecast Ridership in average weekday for a project.
Actual Ridership	Observed Ridership in average weekday for a project.
Project Development Phase	Defined as the planning phase in which the forecast was made. Planning/environmental, engineering/design and funding decision phase.

Field	Definition
Year Forecast Produced	The year the forecast was generated.
Forecast Year	The future year for which the forecast was generated.
Year of Observation	The year that actual ridership was observed. Many projects have multiple observed ridership values. Actual ridership from the year closest to the forecast year is used.
Mode	Primary mode of the transit system. Can be one of Bus, Light Rail, Commuter Rail, Downtown People Mover, Streetcar/Trolley and Urban Heavy or Light Rail.
Number of stops	The number of stops added/served by the project.
First mode	Whether the project introduces first of its kind in the system.
Length	Length of the transit system.
Servicing Central Business District	Whether the project services the central business district.
Service Level	The project's assumed frequency. Actual Value as a percentage of assumed value.
Travel Time	Time to travel from end to end. Actual Value as a percentage of assumed value.
Fare	Project fare per unlinked passenger trips.
Supporting transit systems	Existing transit systems in the service area.

2.3 Method for Quantifying Forecast Accuracy

The accuracy and bias of point forecast is relatively simple to evaluate, although the interpretation is dependent on which metric is used. From the review of past studies, we see two schemes for evaluating forecast performance: as a percentage error and as a ratio. Within those schemes, there is some disagreement as to whether the percentage error should be taken relative over the observed count or over the forecast value, and as to the direction of the sign. Consistent with (Bent Flyvbjerg et al. 2006) and others, we expressed the percent difference in counted traffic from the forecast as:

$$PDF = \frac{Count - Forecast\ Volume}{Forecast\ Volume} * 100\% \qquad Equation\ 1$$

Where PDF is the percent difference from forecast. Negative values indicate that the counted volume was lower than the forecast, and positive values indicate the counted volume was higher than the forecast. It expresses the deviation relative to the forecast, so provides meaningful information when making a forecast. While some authors refer to this expression as Percent Error (PE), we prefer the PDF terminology because it makes the directionality clear.

Whereas PDF measures accuracy for a single project, we are interested in measuring accuracy across a sample of projects. Accuracy is comprised of trueness (lack of bias) and precision, as Figure 3 illustrates (Collaboration for Nondestructive Testing n.d.). In the context of scientific measurement, trueness is the agreement between the average of a large series of measurements and the true value, and precision is the agreement between repeated measurements of the same quantity (ISO 5725-1 1994). These terms do not explain why an outcome occurred: an error does not imply a mistake, and bias does not imply a lack of objectivity.

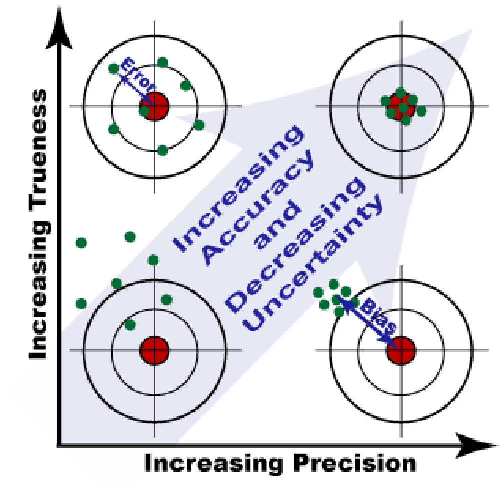

Figure 3 Accuracy and uncertainty terminology (Collaboration for Nondestructive Testing n.d.)

Several differences arose when we translated these terms into the context of traffic and transit ridership forecasting. A simple rendition would take the post-opening count as representative of the true value and the forecast as a measurement. However, a count is itself a measurement subject to substantial error from temporal variation and traffic mix (Ismart 1990; Horowitz et al. 2014). A forecast, on the other hand, is distinct from a measurement because of the time between making a forecast and observing an outcome. In addition, we rarely have repeated traffic forecasts for the same road project, so could not measure precision through repeated measurements. Nonetheless, distinguishing between the components of accuracy is useful. Instead of trueness, we reported the mean and median PDF as measures of the overall deviation. Instead of precision, we reported half the difference between the 5th and 95th percentiles as a measure of the spread of outcomes after adjusting for the average deviation. We separately reported the mean absolute PDF (MAPDF) as a measure of the general accuracy.

To test whether the categorical variables have a statistically significant effect on forecast accuracy, we perform the Analysis of Variance (ANOVA) tests. The hypothesis tested by the one-way ANOVA are:

Null Hypothesis (H_0): The mean PDF across the variables are equal, e.g. the mean PDF for projects opening from 2008 to 2014 is statistically no different from those opening between 2003 to 2007 and these are equal to the overall mean PDF.

Alternate Hypothesis (H_A): At least one mean PDF is different from other groups.

The hypothesis is accepted or rejected by their F-statistics— if the p-value associated with the F is smaller than 0.05, then the null hypothesis is rejected and the alternative hypothesis is supported at a 95% confidence level. If the null hypothesis is rejected, we can conclude that the means of all the groups are not equal. Once the significance is established, the within group variability is tested to explore their effects on the population. Theis variability in the form of pairwise differences in mean is estimated by Tukey's Honestly Significant (HSD) Test.

2.4 Method for estimating factors affecting forecast uncertainty

We begin by defining the conceptual relationship between accuracy and uncertainty. Then we present an econometric framework to measure accuracy and estimate uncertainty windows.

2.4.1 The Conceptual Relationship between Forecast Accuracy and Uncertainty

As Figure 4 illustrates, accuracy and uncertainty are deeply intertwined concepts, especially in the context of forecasting in planning. Accuracy is the closeness of a measurement or estimate to its true value (ISO 5725-1 1994). Uncertainty is the range in which a true value lies with some level of confidence (ISO/IEC Guide 98-3 2008). In forecasting, we treat post-opening traffic count as an observation of the true value with the caveat that the counts themselves are subject to measurement error. Evaluating accuracy is a retrospective activity that accounts for past forecast errors, while expressing uncertainty is a prospective activity that considers possible errors. Because an uncertainty estimate is a "*means of expressing the accuracy of results*" (Collaboration for Nondestructive Testing n.d.), we should consider observations of historical accuracy when estimating uncertainty windows. We propose that the comparison of observed versus forecast traffic for past projects should be used to estimate the range of possible traffic volumes in future forecasts.

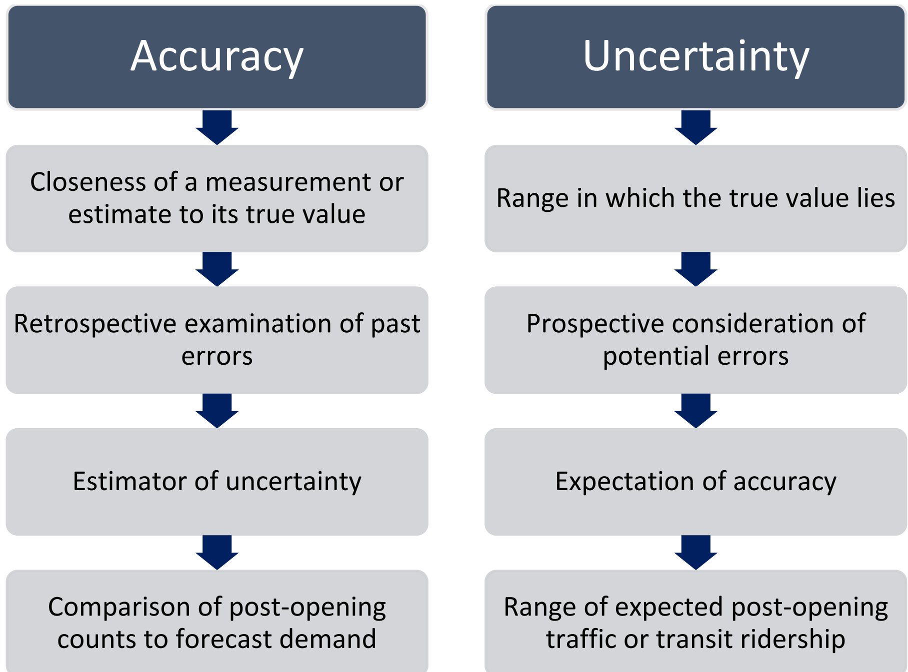

Figure 4: Relationship between traffic forecast accuracy and uncertainty

2.4.2 Theoretical Background of Uncertainty Analysis

Past studies looking at traffic forecast accuracy have identified several factors that contribute to inaccuracy. Methodological weaknesses and the high degree of uncertainty in socio-economic predictions exogenous to the forecasting model have been found to be contributing factors. In addition, researchers have attributed the ramp up period of forecasts, forecast horizon to affect the accuracy as well. Those scholars and critics have offered possible reasons for forecast inaccuracy, including poor data on which forecasts are based, incorrect assumptions about future conditions, limitations of the forecasting methods used, and political motivations that sometimes that cause people to distort forecasts intentionally.

(G. D. Erhardt et al. 2019) analyzes the effect of the above-mentioned variables on traffic forecast accuracy. The study attributed the "*general over-prediction of traffic*" to be contingent on the project type (new construction, on existing roads), roadway characteristics (functional class, traffic volume), forecast characteristics (methods, forecast horizon) and economic forecasts.

Transit ridership has been documented to be affected by a myriad of factors to a varying degree. Internal factors like the amount of service provided, the fare and the reliability and speed of the transit system affect the relative utility of transit compared to the other modes present in the transportation economy. According to TCRP Synthesis 66, there is a lack of transferability of transit ridership forecasts since the models are developed and applied on local bases (Boyle, Board, and Program 2006). Research have shown that external factors such as demographic and socio-economic state of the locality as well as the presence and efficacy of competing choices can also impact the transit ridership. Such relationships are simple to theorize but their sensitivity is dependent on the characteristics of the service area the agencies operate in. For example, the impact of factors affecting transit ridership in a dense metropolitan like New York or Washington, DC will not be the same for a sparsely populated area like Lexington, KY. Therefore, in addition to reference classes in project types and modes, quantifying the uncertainty in transit ridership forecasts need to accommodate locality types as reference classes as well.

Two common methods to address this uncertainty are sensitivity testing and scenario analysis— both involve running a forecasting model using variable inputs or model parameters. In sensitivity testing, the analyst identifies a few key input variables (e.g. population growth, travel time, toll rate etc.), and runs the travel demand model with upper and lower limits in these variables (Kriger, Shiu, and Naylor 2006; Briggs et al. 2012). However, testing variables one at a time does not address the complex interaction among them (Adler et al. 2014). Scenario analysis addresses this limitation by defining thematic scenarios that consider multiple variables together, such as optimistic, pessimistic and most-likely cases (Davidson 2014; Lyons and Davidson 2016; Lyons and Marsden 2019). Forecasters can combine either approach with Monte Carlo simulation, which uses a probability distribution of the input variables and constructs a distribution of outputs (Lemp and Kockelman 2009; Aldrete et al. 2010; de Jong et al. 2007; Manzo, Nielsen, and Prato 2015a; 2015b). Others are exploring ways to efficiently define the scenarios to run (Knaap et al. 2020).

All these approaches rely on some understanding of the distribution for each input variable tested. In practice, "the levels of each variable tested are typically arbitrarily set and *do not correspond to any particular likelihood of occurrence*" (Adler et al. 2014). Alternatively, forecasters can use input variable distributions derived from the historical variance and covariance of those variables (de Jong et al. 2007). While the future probability of an event may differ from its past frequency and forecasts will always contain assumptions, behavioral economists have long recognized *"major deficiencies in the unaided, intuitive judgments of probabilities for uncertain events"* (Kahneman and Tversky 1977). Therefore, it is appropriate to use past data as an aid to understand future distributions.

2.5 Quantile Regression Methodology

We propose an empirical method for estimating uncertainty windows around traffic forecasts that is based in data on the historical accuracy of forecasts. Specifically, we develop quantile regression models of post-opening traffic volumes as a function of forecast traffic volumes and of project attributes. We apply these models to calculate a range of expected traffic volumes for future forecasts, based on the 5th and 95th percentile estimates, as well as the expected median traffic volume. Ascher (1979)'s outsider's approach and Kahneman and Tversky (1977)'s reference class forecasting inspire our approach. Whereas the traditional insider's approach considers possible uncertain events or parameters and builds up to a range, the outsiders view considers a project relative to a statistical distribution of past outcomes from a comparable reference class of projects. For example, in project scheduling, the insider's approach estimates the duration of each task and sums to a total, whereas the outsider's approach looks at the average duration of similar completed projects. Flyvbjerg (2007) recommends the use of reference class forecasting for large infrastructure projects and Schmitt (2016) demonstrates how to use it in transit forecasting. Those wishing to apply it face the challenge of defining the reference class while maintaining an adequate sample size within that class. Our method incorporates variables related to accuracy into the quantile regression models, capturing their effect on the ranges without subdividing the sample.

Whereas most studies focus on reporting descriptive statistics of forecast errors, Odeck and Welde (2017) define and apply a formal econometric framework for evaluating forecast accuracy. The econometric framework is advantageous because it provides a simple, but statistically robust method for estimating the bias. It does so by estimating the following regression:

$$y_i = \alpha + \beta \hat{y}_i + \varepsilon_i \qquad Equation\ 5$$

Where y_i is the actual traffic on project i, $\hat{y}_i$ is the forecast traffic on project i, and ε_i is a random error term. α and β are estimated terms in the regression. The null hypothesis is that the forecasts are unbiased, and in that case the estimated value of α will be 0 and of β will be 1.

It is easy to see how this econometric framework can be extended to test additional segmentation, or additional terms in the regression. For example, either α and β can be segmented by the type of project, the agency conducting the forecast, or the number of years between the forecast and the opening year. This provides a framework from which a wealth of factors can be explored with different levels segmentation depending on the number of observations in each segment.

This research will do so by following the Odeck and Welde (2017) structure, but introducing additional terms as descriptive variables:

$$y_i = \alpha + \beta\hat{y}_i + \gamma X_i + \varepsilon_i \qquad \textit{Equation } 6$$

where X_i is a vector of descriptive variables associated with project i, and γ is a vector of estimated model coefficients associated with those descriptive variables. In this formulation, δ =0 indicates no effect of that term, while positive values would scale up the forecast and negative values would scale down the forecast. The coefficients of categorical variables signify their effect compared to an omitted reference level. For example, consider a model in which α is 0, β is 1 and there is a single descriptive variable, $X_{1,i}$, a binary flag which is 1 if the forecast is for a new road, and 0 for a project on an existing roadway. If δ_1 has a value of -0.1 the expected value would be 10% lower than the forecast. If δ_1 has a value of +0.1 the traffic count would be 10% higher than the forecast.

With the above formulation we can explore the variables associated with higher or lower traffic relative to forecast but can say nothing of the distribution beyond the mean. For example, forecasts with longer time horizons may be no higher or lower on average but may have a wider range of outcomes. Therefore, we extend the above framework to use quantile regression instead of ordinary least square (OLS) regression. Whereas OLS predicts the mean value, quantile regression predicts the values for specific percentiles in the distribution (Cade and Noon 2003). In addition, Quantile Regression Methodology does not assume any parametric distribution (e.g. normal, Poisson etc.) of the random error term in the model, unlike OLS. Zhang and Chen (2019) used quantile regression to quantify the effect of weather on travel time reliability, where an event may have a small effect on the mean value but increase the likelihood of a long delay. In an application analogous to this project, Pereira et al. (2014) used quantile regression to estimate error bounds for real time traffic predictions.

Quantile regression is like linear regression. Instead of computing the standard errors based on the sample mean, the errors are based on a specified quantile (e.g., 10th percentile, 20th percentile, etc.) of the sample. Quantiles, or percentiles, are a cut points that divide a frequency distribution into intervals with the specified probability. For example, the 5th percentile (quantile 0.05) is the value for which there is a 0.05 probability of a value drawn randomly from the distribution being lower than the specified value. At the 95th percentile, there is a 0.05 probability of a value drawn randomly from the distribution being higher than the specified value. Therefore, a range of quantiles can be used to express a range of likely outcomes.

Whereas linear regression would estimate a single α and single β, quantile regression instead estimates one α for each quantile of interest and one β for each quantile of interest. Such a model must be estimated based on historic data—using forecasts that were made in the past for projects that have since opened, such that actual data can be collected. Whereas OLS predicts the mean value, quantile regression predicts the values for specific percentiles in the distribution (Cade and Noon 2003). Quantile regression has been used in transportation in the past for applications such as quantifying the effect of weather on travel time and travel time reliability (Zhang and Chen 2017), where an event may have a limited effect on the mean value but increase the likelihood of a long delay. It has also been used to estimate error bounds for real time traffic predictions (Pereira et al. 2014), an application more analogous to this project.

An example of a quantile regression plot of counted traffic as a function of forecast traffic has been provided in Figure 5.

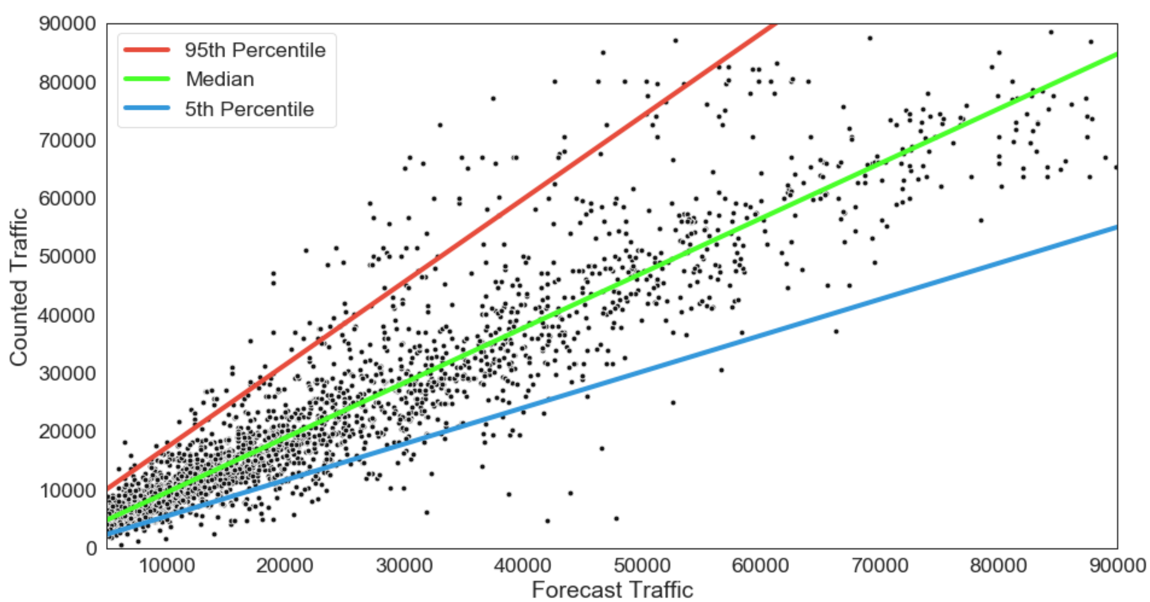

Figure 5: Sample Quantile Regression Plot

For example, assume Figure 5 represents a series of project forecasts and their actual values. Each point represents one project. The forecast ADT was predicted several years prior for the project opening year. The actual ADT was measured after the project opened. Consider that our goal is to predict the actual ADT as a function of the forecast ADT. Estimating such a model using standard linear regression, would result in the line drawn through the middle of the cloud of data. Quantile regression, on the other hand, draws lines along the edges of the cloud, essentially creating an envelope of probabilistic range. For example, if the top line represents the 95th percentile values and the bottom one 5th percentile, we can say that all the points between these two lines contain 90% of the data points. For a particular forecast value, therefore, we can estimate the upper and lower bounds which represent the range of 90% actual observations. Based on historic accuracy, if we have a forecast of X, we would expect that 90% of actual outcomes to fall between this range.

2.6 Factors Affecting the Recent Transit Ridership Decline

The quantile regression method proposed here relies on data that include the full set of deviations occurring in the past, including the travel effects of events such as the 2008 financial crisis and fluctuating gas prices. However, this data-driven approach may be limiting if the future looks discontinuous from the past. For example, the effect of self-driving vehicles may pose a risk to forecasts made for 2040, and outcomes for projects that have already opened cannot clarify that risk. The National Road Traffic Forecasts (NRTF) in the UK recognizes this challenge and addresses them by investigating factors that most influence road traffic and their relation to such unknown and imminent changes in travel behavior (introduction of connected and autonomous vehicles in the network, changes in transportation policy etc.) (Lyons and Marsden 2019). There has been a sudden shift in the way people travel in 2020 because of the COVID-19 pandemic. Stay at home orders were enforced throughout 2020 and the early parts of 2021. There was a surge in unemployment across the world and teleworking was the norm for most of the white-collar industry. Studies show that future aggregate travel behavior is due for a massive change as the economy starts to recover. For quantifying the uncertainty traffic and transit ridership forecasts, this changed travel behavior needs to be addressed.

For transit ridership forecasts, one of the unexpected changes has been the decline in ridership across the US between 2012 and 2018. During this time, bus ridership in the declined 15% and rail ridership declined 3%. While these trends are remarkably consistent across US cities, transit ridership in other countries has increased in the last several years, with the few countries experiencing ridership losses also suffering from poor economic conditions or substantial demographic changes (Freemark, 2019; Miller et al., 2018; UITP, 2017), which the US did not. While gas prices are lower, the US experienced a strong economy, stable demographics and improved transit service over this period, making these ridership losses surprising. For the uncertainty estimation models to be useful to transit agencies forecasting their ridership, the models need to incorporate the factors that have been affecting ridership in recent years as well.

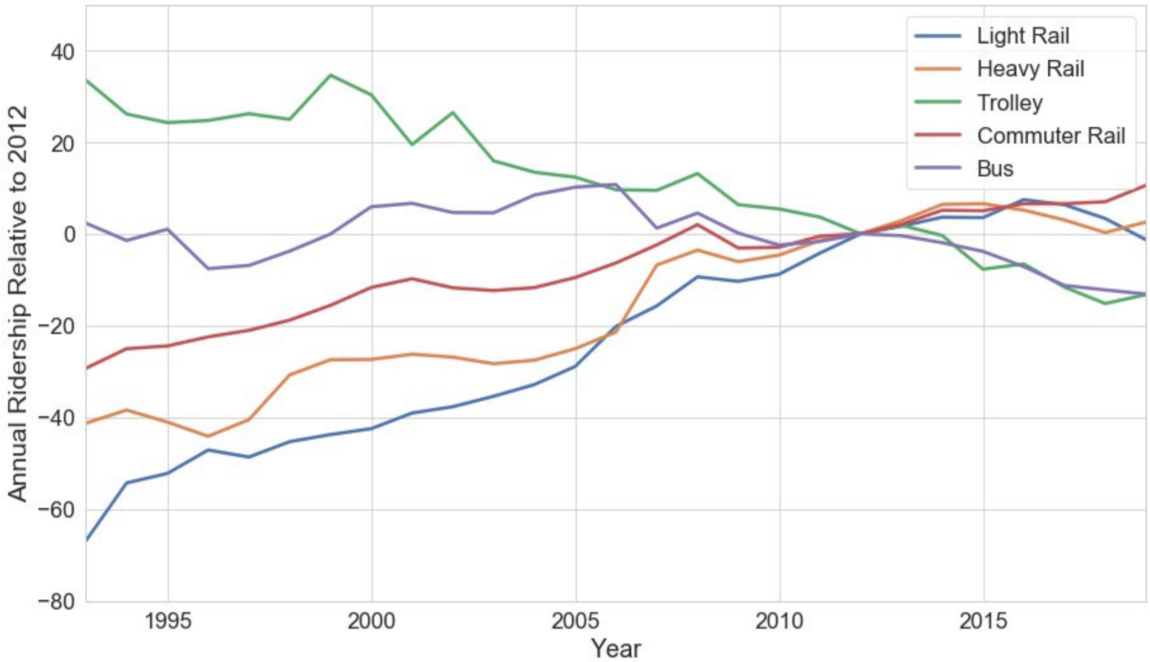

Figure 6 Annual Ridership Change relative to 2012 by Mode (Source APTA Ridership Report)

In order to capture the system-level change in the transit ridership in recent years, this research employs longitudinal models of total bus ridership and rail ridership for 260 Metropolitan Statistical Areas (MSAs) in the US for the period from 2002 through 2018. The model results establish the sensitivity of transit ridership to changes in the descriptive variables (service miles, fares, population, economic condition of the locality, presence of alternate modes of transportation etc.) covering both the period of recent decline and a longer reference period. The variation across both time and space allows for better statistical estimates of the sensitivity to these variables because they may change at different rates in different MSAs. Transit agencies in the United States operate in a wide variety of environments, from small towns to mega regions, where decades of urban planning have shaped the way people travel. The principle of grouping metropolitan areas based on socio-economic and demographic characteristics that affect transit ridership has been applied in (Ederer et al. 2019a). Analysis of transit ridership decline using a similar clustering principle will provide us the context of locality in our analysis of forecast uncertainty.

2.6.1 Data Description

The National Transit Database (NTD) reports time series data of transit profiles and summaries at an agency level. The monthly data (unlinked passenger trips, vehicle revenue miles and vehicle revenue hours and fare revenue) reported by the transit agencies are aggregated by mode, year and by the MSA they serve, replicating (Ederer et al. 2019b) methodology. In our analysis, we only took rail and bus modes that had continuous data available through 2018, in some cases taking the latest consecutive years.

The longitudinal data on the explanatory variables were collected from several different sources: the American Community Survey (ACS) 2005 to 2018, Bureau of Labor Statistics (BLS), Longitudinal Employer-Household Dynamics (LEHD), National Historical Geographic Information Systems (NHGIS), Energy Information Administration (EIA), Bureau of Transportation Statistics (BTS) etc. The detailed list of explanatory variables tested in the econometric models and their sources are presented in Table 4.

Table 4: Description of variables tested to quantify the factors affecting transit ridership decline

Variable	Definition	Source	Unit
Ridership	The ridership variable, Yearly Unlinked Passenger Trips (UPT) times the percentage of people in the Urbanized Area living in the MSA.	NTD	Unlinked Passenger Trips/Year
Service Supply.	Vehicle Revenue Miles, variable describing the Yearly Service Miles of the agency, times the percentage of people in the Urbanized Area living in the MSA. Other related variable used: service supply of competing mode in the MSA.	NTD	Vehicle Revenue Miles/Year
Average Fare	Total Yearly Fare Revenue per UPT value adjusted to 2018 dollars.	NTD	$/Unlinked Passenger Trips
MSA Population and population characteristics	Total Population in the Metro Area. Other variables tested: racial mix, percent of immigrant population, percent of population born in and out of the state of residence, poverty status, and age distribution.	ACS 1-year estimation	#/Year
MSA Employment and employment characteristics	Total Employment in the Metro Area. Other variables tested: unemployment rate.	Bureau of Labor Statistics	#/ Year
Population and Employment Characteristics in the Transit Supportive Density Area in an MSA	Percent of total yearly population living in the Transit Supportive Density, (census tracts that were identified to be transit supportive: Total Population and Employment per acre in 2010>10). Variables tested: wages earned, population percentage employment percentage, education level and racial mix.	Longitudinal Employer-Household Dynamic (LEHD), National Historical Geographic Information System (NHGIS)	Percent of Total Yearly MSA Population
Average Gas Price	Average yearly Gas Price in the MSA, adjusted to 2018 dollars	Energy Information Administration	$ (Inflation adjusted)

Variable	Definition	Source	Unit
Per Capita Median Income	Median Income of individuals adjusted to 2018 dollars. Similar variable tested: median income per household and percent of population earning below 35k yearly.	ACS 1-year estimation	$ (Inflation adjusted)
Car Ownership	Percent of Households with 0 vehicles. Other variables tested: average number of cars per household, number of cars in each household.	ACS 1-year estimation	Percent
Commute Characteristics	Percent of people working from home. Other variables tested: percent of population driving alone, taking carpool, transit or walking, cycling or taking modes other than car.	ACS 1-year estimation	Percent
Years since TNC arrived at the MSA	Number of years since TNCs arrived at that Metro Area. This variable has been segmented to test its effect on Bus and Rail ridership, as well as on MSAs that have transit operating expenses above 300M and New York and MSAs with operating expenses below 300M.	Uber	#
TNC Trips per Person in 2017	Geographic breakdown of total TNC trips in 2017 in the high operating expense MSA group as defined by APTA, and single rates for each in mid and low operating expense cluster divided by the total population. Extrapolated to the other years based on their population and market share as well as the presence of TNC in the MSA for a particular year.	National Household Travel Survey 2017	TNC Trips/person
TNC Revenue	Revenue data obtained from Uber and Lyft SEC report, extrapolated to each MSA where TNC was available for a particular year	Uber and Lyft SEC report	$
Per capita TNC Trips	Number of TNC trips per capita in the metro area, calculated from TNC revenue share by MSA by TNC arrival date and TNC trips per Person. Other variables tested: TNC revenue per capita.	Calculated	TNC Trips/Person

Variable	Definition	Source	Unit
Presence of Bike Share	Presence of Bike Sharing system in the Metro Area. Other variables tested: number of dockless and docked bikes in the MSA.	Bureau of Transportation Statistics	Binary (1 if present, 0 if not)
Presence of Electric Scooter Share	Presence of e-scooter sharing system in the Metro Area. Other variable tested: number of e-scooters in the MSA.	Bureau of Transportation Statistics	Binary (1 if present, 0 if not)
Mean Distance Between Failure	Vehicle Revenue Miles/Mechanical System Failures for Revenue Vehicles. Measure of service quality and reliability. Categorized by other mechanical failures or major mechanical failures.	NTD	Miles
Maintenance and Restructure	Binary variable for identifying whether any maintenance or network restructuring works were conducted in a particular year.	TCRP Report 140	Binary (1 if present, 0 if not)
Travel Time Index	The ratio of travel time in the peak period to the travel time at free-flow conditions.	Urban Mobility Report (2019) (Schrank, Eisele, and Lomax 2019)	Value
Network Restructuring	Restructured bus network, changed routes and service allocation.	TCRP Synthesis 140 (Byala et al. 2019)	Binary (1 if present, 0 if not)
Major Maintenance	Safety incidents in 2015 and 2016 on the Washington Metro led to line closures and major maintenance work in the following years, with disruptions lasting from late 2015 to early 2018.	Major news outlets	Binary (1 if present, 0 if not)

2.6.2 Method

The sensitivity of transit ridership to changes in the descriptive variables is established through a longitudinal analysis of mode level transit ridership. Such relations vary across the metro areas as well as over time and are estimated through a Panel Ordinary Least Squared (OLS) Model.

Transit ridership is in essence a demand-supply problem. The relative utility of transit compared to the other modes depends on the supply (frequency, density of stops, accessibility, proximity to attractions etc.) as well as the fare. This supply is in turn dependent on the ridership—the more people using the service, the more the agencies are prompted to increase their service. This endogeneity violates the basic assumption of regression. In addition, it is not possible to include every factor in the analyses: as described in the previous section, we didn't consider several variables in our dataset because of their unavailability. These omitted variables are also likely to interact with the other variables in the model, producing biased estimation. Assuming unobserved factors at each MSA that might simultaneously affect the ridership and the demographic variables do not change over time, we consider Fixed Effect in our model estimation. Fixed effect models avoid the unobserved heterogeneity and endogeneity biases by using each individual entity as their own control in time. Fixed effects models control for the effects of time-invariant variables with time-invariant effects. This is true whether the variable is explicitly measured or not. If y_{it} is the total ridership, or UPT, for Metropolitan Statistical Area i at year t, and x_{it} are the explanatory variables, the standard format of fixed-effect Panel OLS is:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \sum_{j=1}^{n-1} d_j \alpha_j + \varepsilon_{it} \qquad Equation\ 2$$

Where d_j is a dummy variable equal to 1 for MSA j and 0 for the others. There are n-1 dummy variables, one for each MSA except the last one whose fixed-effect is merged with the constant term. α_j is the fixed effect for MSA j.

The model itself can take different forms. Since ridership is essentially a count data, it is skewed and its variance increases with their mean. Skewed data can be transformed using the natural logarithm as long as they have constant variance to the mean. In our analysis we estimated a mixed log-log and log-linear model noting the non-linear relationship of ridership with k dependent variables as well as the skewness of the data.

$$\log(y_{it}) = \beta_0 + \beta_1 \log(x_{kt}) + \beta_2 x_{(i-k)t} + \sum_{j=1}^{n-1} d_j \alpha_j + \varepsilon_{it} \qquad Equation\ 3$$

Taking the exponent,

$$y_{it} = e^{\beta_0} * x_{kt}^{\beta_1} * e^{\beta_2 x_{(i-k)t}} + u + \in \qquad Equation\ 4$$

The coefficients of the regression represent elasticity of ridership against the log transformed explanatory variables. For the non-transformed variables, each unit increase in X multiplies the expected value of Y by e^{β}.

2.7 Summary of the Research Approach

From the discussion in this chapter, the research approach can be summarized into the following three tasks which correspond to the three objectives of this dissertation:

- Establish empirical evidence of uncertainty in traffic and transit ridership forecasts and factors affecting it through categorical exploration of observation against forecast.

- Identify factors affecting transit ridership in recent years and quantify their effects on transit ridership forecast accuracy.
- Estimate quantile regression models for both traffic and transit ridership forecasts incorporating the factors identified previously.

Chapter 3 THE ACCURACY OF TRAFFIC FORECASTS

Conducted as part of the National Cooperative Highway Research Program (NCHRP) Project 08-110: Traffic Forecast Accuracy Assessment Research, this chapter gives an overview of the accuracy of road traffic forecasts. The analysis is based on a database of 1291 open projects from 6 states and 4 European countries. The chapter answers the following questions:

- What is the state of accuracy in road traffic forecasts over time?
- What factors affect the traffic forecast accuracy?
- Are there any noticeable effects of the Great Recession on traffic forecast accuracy?

We found measured traffic is on average 6% lower than forecast volumes, with a mean absolute deviation of 17% from the forecast. Higher volume roads, higher functional classes, shorter time spans, and the use of travel models all improved accuracy. Unemployment rates also affected accuracy—traffic would be 1% greater than forecast on average, rather than 6% lower, if we adjust for higher unemployment during the post-recession years (2008 to 2014). Forecast accuracy was not consistent over time: more recent forecasts were more accurate, and the mean deviation changed direction. Traffic on projects that opened from the 1980s through early 2000s was higher on average than forecast, while traffic on more recent projects was lower on average than forecast. This chapter has been adapted from the following published paper:

Hoque, J.M., Erhardt, G.D., Schmitt, D., Chen, M., Chaudhary, A., Wachs, M. and Souleyrette, R.R., 2021. The changing accuracy of traffic forecasts. *Transportation*, pp.1-22.

3.1 Introduction

Accuracy of traffic forecasts have been a point of interest among researchers for several decades now, although most of them have been focused on toll roads. The inspiration seems to be from the fact that toll road forecasts have a bearing on investor expectations and that is why their accuracy is more important. As an evidence to this, the Australia Government (2012) cited ''*inaccurate and over-optimistic''* traffic forecasts as a threat to investor confidence. Three lawsuits now underway challenge the forecasts for toll road traffic that subsequently came in significantly under projections (Bain 2013). Most studies on the accuracy of traffic forecasts found that post-opening traffic on free roads was 3-11% higher on average than forecast (Nicolaisen and Driscoll 2014). The most diverse samples were a study of 183 large road projects (both tolled and free) in 14 countries that found counted traffic was an average of 9.5% higher than forecast (Bent Flyvbjerg, Holm, and Buhl 2005), and a study of 146 road projects in Denmark, Norway, Sweden and the United Kingdom (UK) that found counted traffic was on average 11% higher than forecast (Nicolaisen 2012). Most other studies of road forecast accuracy analyzed a single state or country. Welde and Odeck (2011) found counted traffic on Norwegian road projects 19% higher than forecast. Parthasarathi and Levinson (2010) examined the accuracy of traffic forecasts for one city in Minnesota and found forecasts underestimating counted traffic, especially for high volume roads. Buck and Sillence

(2014) evaluated 131 forecasts in Wisconsin and determined that the mean absolute percent difference between forecast and counted traffic was 16%. Giaimo and Byram (2013) analyzed over 2,000 traffic forecasts for road segments in Ohio produced between 2000 and 2012 and found that counts were slightly lower than forecast, but the difference was within the standard error of traffic count data. In contrast to most studies, Miller et al. (2016) reported that counts were *lower* than forecast, with a median percent error of 31% for 39 road projects in Virginia. While the average difference for free roads was in the opposite direction to and smaller in magnitude than for transit and toll roads, each study of the topic showed substantial forecast inaccuracies.

The constraints in data availability as explained in the previous chapter makes coming to a general conclusion about the existence of systematic bias in forecast as a whole difficult. In this study, we assembled those data, compiling a database of forecast traffic and post-opening traffic counts for 2,611 unique projects, and for 16,697 road segments in six states in the United States (US) and four European countries. Some of the projects had not yet opened; some of the segments did not have traffic count data associated with them, and others did not pass the quality control checks for inclusion in statistical analysis. While we retained all records for future use, we based our analysis on a subset of 1,291 projects and 3,912 segments. We used those data to assess the accuracy of the forecasts, and identify the factors related to better or worse accuracy. The data points span from 1960 to 2017 and do not consider the effects of reduced travel due to COVID-19. The results provide insights into the degree of confidence that planners and policy makers can expect from traffic forecasts and suggests that we should view forecasts as a range of possible outcomes rather than a single expected outcome.

3.2 Methods and Data

Many others have conducted case studies of particular projects, but here I present a statistical analysis of a large sample of projects, and therefore constitutes a large-N study (Bent Flyvbjerg, Holm, and Buhl 2005). Large-N analysis aims to determine how close the forecasts are to observed volumes (Miller et al. 2016). Consistent with (Bent Flyvbjerg et al. 2006) and others, we expressed the percent difference in counted traffic from the forecast as:

$$PDF = \frac{Count - Forecast\ Volume}{Forecast\ Volume} * 100\% \qquad Equation\ 1$$

Where PDF is the percent difference from forecast. Negative values indicate that the counted volume was lower than the forecast, and positive values indicate the counted volume was higher than the forecast. It expresses the deviation relative to the forecast, so provides meaningful information when making a forecast. While some authors refer to this expression as Percent Error (PE), we prefer the PDF terminology because it makes the directionality clear.

Whereas PDF measures accuracy for a single project, we are interested in measuring accuracy across a sample of projects. We reported the mean and median PDF as measures of the overall deviation. Instead of precision, we reported half the difference between the 5^{th} and 95^{th} percentiles as a measure of the spread of outcomes after adjusting for the average deviation. We separately reported the mean absolute PDF (MAPDF) as a measure of the general accuracy.

As Flyvbjerg (2005) recommends, we evaluated opening-year conditions. We defined the opening-year as the first post-opening year with traffic count data available. If we had multiple forecasts for a single project (such as opening-year and design-year forecasts, usually 20 years after project opening), we used the forecast closest to the opening-year. To make the comparison in the same year, we held the counts constant and scaled the forecast to the year of the count using the growth rate implied by opening and design year forecasts, and a standard growth rate of 1.5% if they were unavailable.

Projects did not always open in the year anticipated. This happened if a project was delayed, if a forecast was for an alternative design that was not built, or if funding priorities changed. We usually knew when delays occurred for large projects. For smaller projects we could not always determine when and if construction finished because DOTs do not necessarily link forecast records to construction records. Where we could not verify the project completion date (for 488 projects out of the 1291 in our analysis), we assumed that maintenance, minor construction or low risk projects were completed within one year of planned opening, and that major construction projects took two years beyond that. This assumption reduced the risk of including counts collected prior to the project opening. Because most projects took place on existing facilities, pre-opening counts are often available but may be affected by construction activity.

When comparing forecasts and counts, we compared the Average Daily Traffic (ADT), although its exact definition depended upon the source. Some agencies provided data as Average Annual Daily Traffic (AADT), some as Average Weekday Daily Traffic (AWDT), and some as typical weekday traffic, which usually was for non-holiday weeks with school in session. The units were not always clear in the data, so they may vary between agencies, but we assumed consistency between forecasts and counts within an agency.

Often, the forecasts included estimates of traffic on multiple road segments. These estimates were likely correlated, such as for different directions of flow on the same road or for two road segments aligned end-to-end. Rather than retain separate observations for each segment, we aggregated them to a single project-level observation by averaging the forecast and observed traffic volume for all segments with available forecast and count data.

3.3 Overall Distribution of Forecast Accuracy

Figure 7 shows the overall PDF distribution, replicated here from NCHRP 934 (Erhardt et al. 2020) with permission from TRB, which reveals that counted traffic was lower than forecast on average. About 68.5% of projects had traffic lower than forecast. The mean PDF was -5.6% and the mean absolute PDF was 17.3%. The 5^{th} percentile PDF was -37.6% and the 95^{th} percentile PDF was +36.9%. The average difference was opposite in direction from the results of most previous studies of toll-free road traffic forecasts. This difference reflects the composition of the sampled projects, whether by location, type of project, year, or some other factor. We explore how the accuracy relates to such factors in the rest of this section.

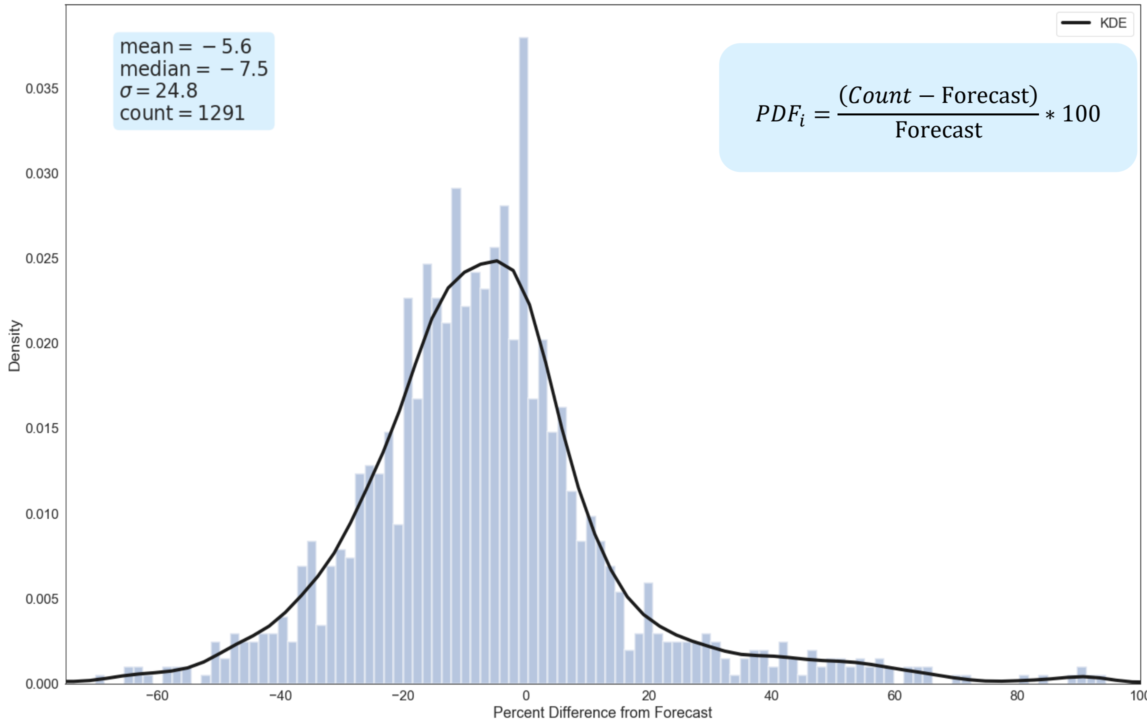

Figure 7: Distribution of Percent Difference from Forecast (Erhardt et al. 2020)

NCHRP Report 255 (Pedersen and Samdahl 1982) provided recommendations on the maximum desirable deviation of a traffic assignment model from base year traffic counts. According to this guidance, assignment deviation should not result in a design deviation of more than one highway travel lane. NCHRP Report 765 described this as the "*half-lane rule*" and extended it by considering the approximate error in traffic counts, in the expectation that an assignment model would not reasonably have less error than traffic counts (Horowitz et al. 2014). In Figure 8, we plotted the absolute PDF against the forecast volume, and overlaid the maximum desirable deviation and expected deviation of traffic counts. We found that 83.8% of forecasts fell within the maximum desirable deviation, and 46.5% of forecasts had less deviation than expected of traffic counts.

To more explicitly test forecasts against the half-lane rule, we calculated the number of lanes required for forecast traffic and counted traffic on each road segment, assuming the same Level of Service. Some 36 segments out of 3912 (1.0%) would have required an additional lane to allow the traffic to flow at the forecast level of service (LOS). Conversely, forecasts for 158 links (4.2%) over-estimated the traffic by an amount such that they could provide adequate service with fewer lanes per direction; 92 of those links were interstate highways, 64 were principal arterials and the rest were minor arterials.

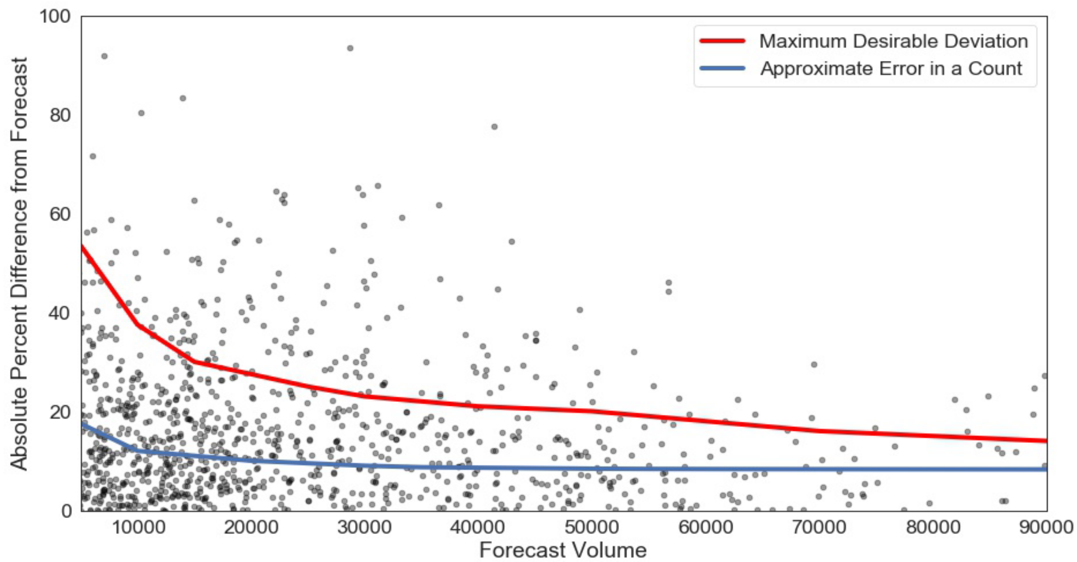

Figure 8 Absolute Percent Difference from Forecast as a function of forecast volume

3.4 Categorical Assessment of Traffic Forecast Accuracy

Table 5 presents the statistical measures of available categorical variables. We discuss the values below.

Jurisdiction: We observe that some agencies have more accurate forecasts than others, although the sample sizes are small for some, and we do not know whether this accuracy is due to better forecasting techniques or a different mix of projects. We noted previously that Agency E recorded nearly every forecast they made since the early 2000s, comprising 44% of our sample. The Agency E projects have a lower absolute deviation (MAPDF of 13.7% compared to 20.1% for the rest), which may relate to including more routine projects. However, their average deviation is more negative (mean PDF of -9.5% against -2.7%), which may be because many of these projects opened in the wake of the Great Recession.

Functional Class: The results show that forecasts were more accurate on higher functional class facilities. Higher functional class roads carry more traffic than other road classes, so a similar absolute deviation is associated with a smaller percent deviation. In addition, smaller facilities may be more affected by zone size and network coding details where all traffic from a traffic analysis zone may enter the road network at one location, leading to uneven traffic assignment outputs.

Area Type: The results show little difference between the accuracy of forecasts in rural or mostly rural counties versus those in urban counties.

County Population Growth: We further grouped projects based on whether they were in counties with growing, stable or declining population between the start year and the opening year. Counted traffic in counties experiencing more than 1% growth was about 12.8% less than forecast on average, compared to 7.9% and 8.6% less in counties with declining or stable population. This result suggests that when a large share of the forecast traffic is due to expected population growth, as might be expected in a growing county, there is a risk that the traffic growth does not materialize.

Table 5: Percent Difference from Forecast by Category

	Obser-vations	MAPDF	Mean	Median	5th Percentile	95th Percentile	95th – 5th Percentile / 2
Overall Distribution							
Project Level	1,291	17.3%	-5.6%	-7.5%	-37.6%	36.9%	37.3%
Jurisdiction							
Under Agency E	562	13.7%	-9.5%	-8.5%	-35.5%	12.1%	23.8%
Projects by the rest	729	20.1%	-2.7%	-6.2%	-39.9%	50.2%	45.0%
Under Agency A	381	17.4	-9.1	-10.3	-42.3	22.0	32.2
Under Agency B	1	6.2	-6.2	-6.2	-6.2	-6.2	0.0
Under Agency C	3	7.8	7.8	5.7	2.8	14.2	5.7
Under Agency D	99	31.9	11.7	8.8	-46.5	81.2	63.9
Under Agency F	132	15.9	-10.8	-10.4	-41.3	11.7	26.5
European Project	113	24.1	15.6	5.8	-23.0	66.9	45.0
Functional Class							
Interstate or Limited Access Facility	187	11.2%	-8.7%	-6.8%	-30.5%	7.2%	18.9%
Principal Arterial	403	14.7%	-7.9%	-8.8%	-33.8%	19.3%	26.6%
Minor Arterial	186	16.7%	-7.5%	-8.0%	-36.1%	20.9%	28.5%
Major Collector	145	17.3%	-12.5%	-10.7%	-47.1%	12.1%	29.6%
Minor Collector	10	21.5%	-18.4%	-19.5%	-39.0%	6.5%	22.8%
Unknown	360	23.5%	2.6%	-3.4%	-41.1%	58.8%	50.0%
Area Type							
Rural or Mostly Rural	367	16.7%	-7.8%	-8.8%	-36.1%	24.8%	30.5%
Urban	811	16.6%	-7.6%	-8.4%	-39.1%	28.6%	33.9%
Unknown	113	24.1%	15.6%	5.8%	-23.0%	66.9%	45.0%
County Population Growth between Start Year and Year of Observation							
Declining (<-1% growth)	383	16.1%	-7.9%	-8.5%	-36.0%	19.9%	27.9%
Growing (>+1% growth)	94	14.8%	-12.8%	-14.2%	-28.0%	4.5%	16.2%
Stable (-1% to +1% growth)	509	13.6%	-8.6%	-7.7%	-36.7%	13.9%	25.3%
Unknown	305	25.8%	4.5%	-2.6%	-42.8%	62.6%	52.7%

	Obser-vations	MAPDF	Mean	Median	5th Percentile	95th Percentile	95th – 5th Percentile / 2
Project Type							
Repaving/Resurfacing	618	14.5%	-9.2%	-8.7%	-37.0%	14.1%	25.6%
Capacity Expansion	281	20.2%	1.3%	-3.4%	-34.6%	52.0%	43.3%
New Road	28	10.6%	-9.2%	-8.8%	-19.3%	3.8%	11.6%
Unknown	364	20.4%	-4.6%	-7.6%	-44.0%	46.0%	45.0%
Forecast Method							
Traffic Count Trend	252	22.2%	-0.1%	-5.2%	-39.3%	55.1%	47.2%
Population Growth Rate	7	11.3%	-2.2%	-0.3%	-16.4%	13.9%	15.2%
Travel Demand Model	179	16.9%	-8.4%	-9.7%	-44.9%	27.2%	36.1%
Professional Judgement	177	17.8%	-11.8%	-11.9%	-43.1%	18.5%	30.8%
Unknown	676	15.5%	-5.4%	-6.4%	-34.4%	29.5%	32.0%
Agency Type							
State Department of Transportation	489	21.5%	-0.9%	-5.6%	-41.4%	54.3%	47.9%
Metropolitan Planning Organization	2	6.9%	-6.9%	-6.9%	-7.4%	-6.3%	0.6%
Consultant	237	17.4%	-6.4%	-8.2%	-35.9%	31.4%	33.7%
Unknown	563	13.7%	-9.4%	-8.4%	-35.5%	12.1%	23.8%
Start Year (When the forecast was produced)							
Before 1990	139	32.1%	16.7%	13.9%	-40.2%	88.8%	64.5%
1991 to 2002	123	19.2%	0.0%	-3.4%	-33.3%	45.4%	39.4%
2003 to 2007	465	17.2%	-10.5%	-11.7%	-37.0%	19.5%	28.2%
2008 to 2014	564	13.3%	-8.3%	-6.7%	-38.7%	12.5%	25.6%
Opening Year							
Before 1990	77	28.9%	8.4%	6.1%	-46.8%	89.0%	67.9%
1991 to 2002	49	33.9%	26.3%	31.5%	-19.5%	63.4%	41.4%
2003 to 2007	168	16.5%	-2.3%	-3.3%	-34.1%	40.5%	37.3%
2008 to 2014	879	15.5%	-8.2%	-8.8%	-36.0%	18.3%	27.2%
After 2014	118	17.0%	-13.4%	-12.3%	-45.1%	11.7%	28.4%

	Obser-vations	MAPDF	Mean	Median	5th Percentile	95th Percentile	95th – 5th Percentile / 2
Time Span (Years)							
0	53	12.4%	-7.0%	-5.6%	-30.0%	12.2%	21.1%
1	228	14.4%	-6.0%	-6.7%	-35.4%	26.2%	30.8%
2	345	15.4%	-7.3%	-7.6%	-39.6%	21.2%	30.4%
3	264	17.5%	-7.8%	-9.5%	-36.1%	29.5%	32.8%
4	139	16.0%	-9.9%	-12.0%	-35.4%	20.6%	28.0%
5+	262	23.7%	1.7%	-3.4%	-42.4%	58.7%	50.6%

Project Type: The traffic on routine maintenance projects (resurfacing or repaving) was on an average lower than forecast. Capacity expansion projects had average counts slightly exceeding forecasts and were less accurate. The difference could reflect capacity expansion projects generating more induced traffic. Forecasts for the construction of new roads were more accurate than forecasts on existing roads, but the sample size was small.

Forecast Method: A Large-N analysis such as this offers the potential to assess the performance of tools available to forecasters, although we were limited to those recorded in the data. Regional travel demand models produced more accurate forecasts than traffic count trends. Some forecasters used professional judgment to combine count trends and volume from a demand model. The resulting forecasts were almost as accurate as those based on models alone, suggesting that considering count trends worsened rather than improved traffic forecasts. We do not know the forecast method for about half the projects with a large percentage (562 out of 676 projects) of those in the jurisdiction of Agency E, which did not record that information.

Agency Type: Relative to state DOTs, consultants produced forecasts with a more negative mean difference, but a smaller spread. Projects with an unknown agency type have a smaller spread than either and almost all of these projects are under the jurisdiction of Agency E (562 out of 563). We do not know whether the differences between consultant- and DOT-prepared forecasts are meaningful or if they instead relate to practices that vary across jurisdictions.

Time Span: We defined the time span as the number of years between the start year and the year of count. Forecasts with a span of 5+ years were less accurate, and counts were lower on average than forecasts. The greater the number of years between forecast production and traffic count, the larger the opportunity for changes to have occurred in the economy, land use patterns, fuel prices, and other factors that influence travel. These are all variables that are difficult to predict, but their effects are evident. This finding is consistent with findings by Bain (2009) who concluded that longer-term forecasts are critically dependent on macro-economic projections.

For projects that opened in 2003 or later, traffic was on average lower than forecast; for projects that opened before 2003, traffic counts were on average higher than forecast. Overall, more recent traffic forecasts were more accurate, as measured by the mean absolute PDF. The ANOVA test results and the subsequent Tukey HSD test to compare the differences in mean PDF (visualized in Figure 9 and Figure 10) suggest that there are statistically significant differences in mean PDF across the projects by their forecast production and observation year. In Table 6 we present the results of the Tukey HSD tests for pairwise comparison.

Table 6: Tukey Pairwise Comparison P-Values of Mean PDF

Start Year				
	1991 to 2002	**2003 to 2007**	**2008 to 2014**	
Before 1990	0.001	0.001	0.001	
1991 to 2002		0.001	0.002	
2003 to 2007			0.422	
Opening Year				
	1991 to 2002	**2003 to 2007**	**2008 to 2014**	**After 2014**
Before 1990	0.001	0.0092	0.001	0.001
1991 to 2002		0.001	0.001	0.001
2003 to 2007			0.0231	0.001
2008 to 2014				0.1626

Highlighted cells depict statistical insignificance, i.e. cannot reject null hypothesis

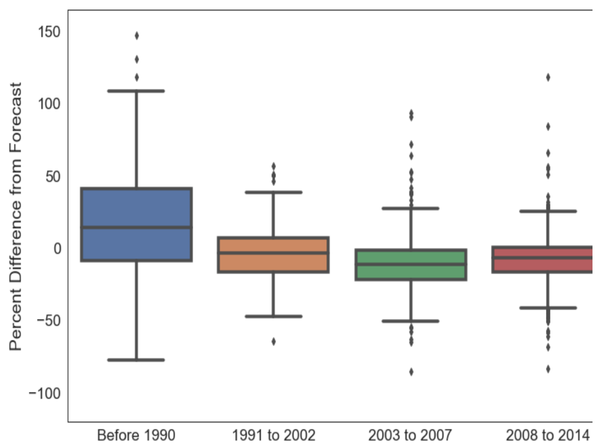

Figure 9: PDF by the Start Year

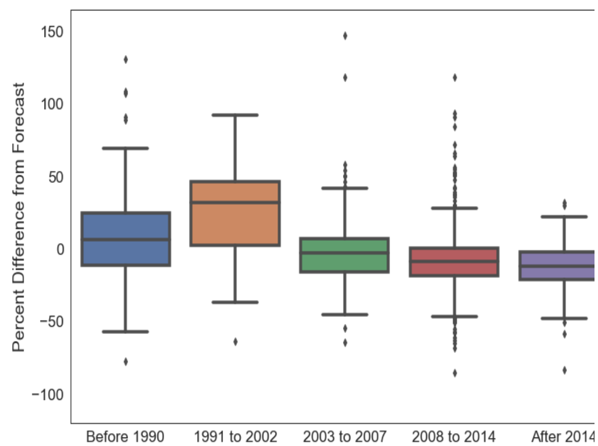

Figure 10: PDF by the Project Opening Year

Unemployment Rate: Economic conditions that differ from expectations can lead to forecast inaccuracy (Anam, Miller, and Amanin 2020). We measured this by examining both the state/country level unemployment rate in the opening year, and the change in unemployment rate from the start year. When the opening year unemployment rates were less than 5%, counted traffic was on average higher than forecast, and when unemployment rates were higher, traffic was lower than forecast. Higher employment rates lead to more traffic as more people commute to and from work. This result highlights the importance of good economic forecasts. Figure 11 and Figure 12 present the box-and-whiskers plot showing their categorical means.

Table 7: Percent Difference from Forecast by Unemployment Rates

	Obser-vations	MAPDF	Mean	Median	5th Percentile	95th Percentile	95th – 5th Percentile / 2
Unemployment Rate in the Opening Year (Percentage)							
0% to 3%	4	19.4%	16.7%	13.1%	-3.2%	41.8%	22.5%
3% to 5%	229	22.9%	2.1%	-2.8%	-40.2%	55.8%	48.0%
5% to 7%	371	16.1%	-7.4%	-7.7%	-39.7%	26.9%	33.3%
7% to 8%	128	17.3%	-7.1%	-6.4%	-43.2%	26.1%	34.7%
8% to 9%	168	17.1%	-5.4%	-7.5%	-33.3%	35.1%	34.2%
9% to 10%	35	18.2%	-5.1%	-11.2%	-28.1%	39.1%	33.6%
More than 10%	356	14.9%	-8.7%	-9.6%	-34.4%	19.6%	27.0%
Unemployment Rate in the Start Year (Percentage)							
0% to 3%	4	18.1%	16.4%	11.3%	-2.1%	41.8%	22.0%
3% to 5%	273	17.4%	-8.1%	-9.3%	-36.1%	26.7%	31.4%
5% to 7%	545	19.9%	-6.7%	-10.5%	-40.2%	36.0%	38.1%
7% to 8%	87	16.8%	-0.4%	-0.1%	-42.7%	42.9%	42.8%
8% to 9%	129	15.0%	-0.5%	-1.9%	-36.1%	54.6%	45.4%
9% to 10%	51	16.8%	-4.7%	-5.8%	-37.5%	31.5%	34.5%
More than 10%	202	11.9%	-5.4%	-5.6%	-26.9%	15.3%	21.1%
Change in Unemployment Rate between Start Year and Opening Year							
Decrease in Unemployment Rate							
	459	18.7%	-1.3%	-4.3%			
-8% to -4%	101	14.7%	-5.9%	-6.1%	-34.6%	31.4%	33.0%
-4% to -2%	136	19.2%	4.4%	-0.7%	-30.6%	54.6%	42.6%
-2% to 0%	367	17.6%	-4.3%	-6.2%	-38.8%	36.6%	37.7%

Increase in Unemployment Rate							
	832	16.5%	-7.9%	-9.3%			
0% to 2%	263	16.8%	-6.0%	-6.3%	-40.6%	30.6%	35.6%
2% to 4%	217	17.1%	-8.0%	-8.6%	-36.1%	26.7%	31.4%
4% to 8%	207	17.5%	-11.5%	-13.9%	-35.7%	19.7%	27.7%

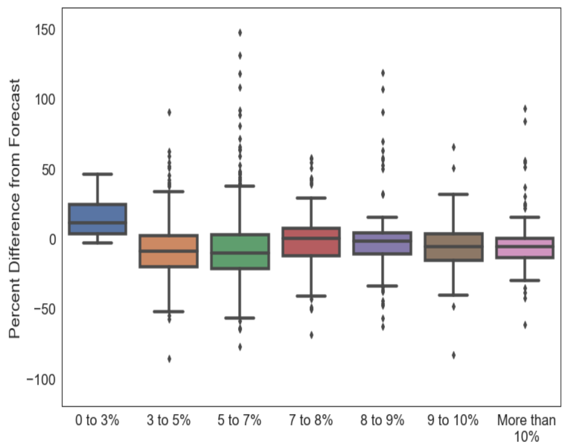

Figure 11: PDF by Start Year Unemployment Rate

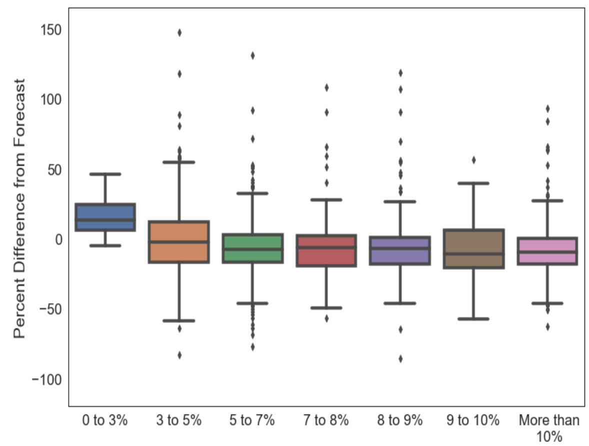

Figure 12: PDF by Opening Year Unemployment Rate

3.5 Effect of the Great Recession on Traffic Forecast Accuracy

The goal of ensuring that transportation funding dollars are being invested wisely needs minimizing the errors by understanding the sources and improving the future modeling practices and forecasting application. This reality has become even more apparent in the wake of the COVID-19 pandemic. With rising unemployment rates and the new normal of working from home and social distancing, travel behavior has changed drastically across the world. Demand for transportation infrastructure projects opening during this time are much likely to be very different from what has been forecasted. In this context, it is important to evaluate the effect of major economic disturbances in the past on the accuracy of traffic forecasts.

Because a large share of projects in our sample opened during or shortly after the Great Recession, we considered how this unexpected event may affect the accuracy of forecasts. To do so, we measured the accuracy of the same traffic forecasts against a counterfactual world in which the Great Recession did not occur. We did this by holding the forecasts constant and adjusting the traffic counts to offset the high unemployment rates observed from 2008 through 2014. Previous work estimated that median post-opening traffic volumes decrease 3% for each percentage point increase in the unemployment rate (Erhardt et al. 2020). We applied this rate to the difference between the opening-year unemployment rate and the pre-recession (2007) unemployment rate for the same county. This process results in adjusted counts that are higher than the true counts for the six-year period in which unemployment exceeded its pre-recession levels. Then we compared the forecasts to the adjusted counts, as Table 8 shows.

For projects opening during the 2008 through 2014 period, post-opening counts are on average 8.2% lower than forecast. However, the recession-adjusted counts are 1.9% higher than forecast. When considering all projects, counts are on average 5.6% lower than forecast, but recession-adjusted counts are 1.3% higher than forecast. Figure 13 shows the effect of this adjustment visually, with the count adjustment shifting the distribution to the right and also spreading it out, as observed in the larger difference between the 5th and 95th percentiles. From these results we conclude that the Great Recession was the major cause of the observed shift, but that other factors cause random deviations resulting in the observed spread.

Table 8: Comparison of descriptive statistics before and after unemployment adjustments

	Obser-vations	**MAPDF**	**Mean**	**Median**	**5th Percentile**	**95th Percentile**	**95th – 5th Percentile / 2**

Projects opening between 2008 and 2014							
Original Sample	879	15.5%	-8.2%	-8.8%	-36%	18.3%	27.2%
Adjusted Sample	879	15.6%	1.9%	1.1%	-30.6%	35.2%	32.9%
All projects							
Original Sample	1291	17.3%	-5.6%	-7.5%	-37.6%	36.9%	37.3%
Adjusted Sample	1291	17.3%	1.3%	-0.4%	-35.4%	42.7%	39.1%

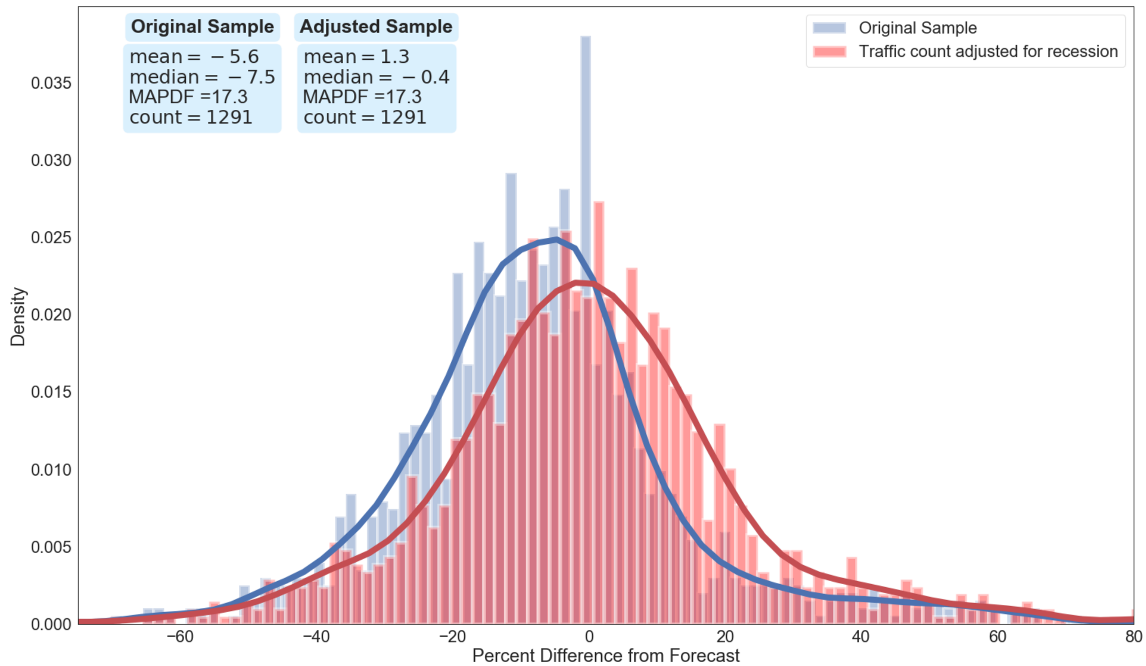

Figure 13: Distribution of Percent Difference from Forecast Adjusting for Great Recession

Year: For projects that opened in 2003 or later, traffic was on average lower than forecast; for projects that opened before 2003, traffic counts were on average higher than forecast. The negative deviation from forecasts in older projects aligns with most previous literature on toll-free road projects (Bent Flyvbjerg, Holm, and Buhl 2005; Parthasarathi and Levinson 2010; Welde and Odeck 2011; Nicolaisen 2012), but it is interesting that the average difference changes direction for more recent projects. Overall, more recent traffic forecasts were more accurate, as measured by the mean absolute PDF.

Several factors could explain these changes. First, better data and improved forecasting methods may have led to more accurate forecasts. Second, the mix of projects in our data may have driven the change, such as the relative frequency of small versus large projects. We would expect non-capacity increasing projects to generate less induced demand than capacity increasing projects, so if induced demand were an important factor in traffic forecast accuracy, then the project mix matters. Third, vehicle miles traveled (VMT) per capita grew rapidly in the 1980s and 1990s. In the 2000s, this trend leveled off and declined, before subsequently rebounding in about 2013. Traffic forecasts might not adequately capture these macro-trends, which appear to be driven largely by gross domestic product (GDP) per capita and fuel price (Bastian, Börjesson, and Eliasson 2016), with possible contributions from other factors such as discount air travel and the substitution of better information and communications technology for travel (G. D. Erhardt 2017).

To further consider these possibilities, we plotted the PDF by opening year alongside the vehicle miles traveled (VMT) per capita in the United States (Source: Davis 2019) in Figure 5. While our data included projects outside the United States, similar VMT trends were observed in Europe (Bastian, Börjesson, and Eliasson 2016). To minimize the impact of changing project types, we excluded repaving projects from Figure 14. Each point represents a single project, and the blue line is a 5-year rolling average of PDF. The figure shows noticeable correlation between PDF and VMT per capita. While VMT per capita was increasing, counted traffic volumes were higher than forecast, but after VMT per capita peaks, the opposite is true. This relationship suggests that traffic forecasts may not have fully captured the factors driving aggregate VMT trends. This relationship between traffic forecast accuracy, aggregate VMT trends and related factors, such as fuel price and economic growth, warrants further investigation.

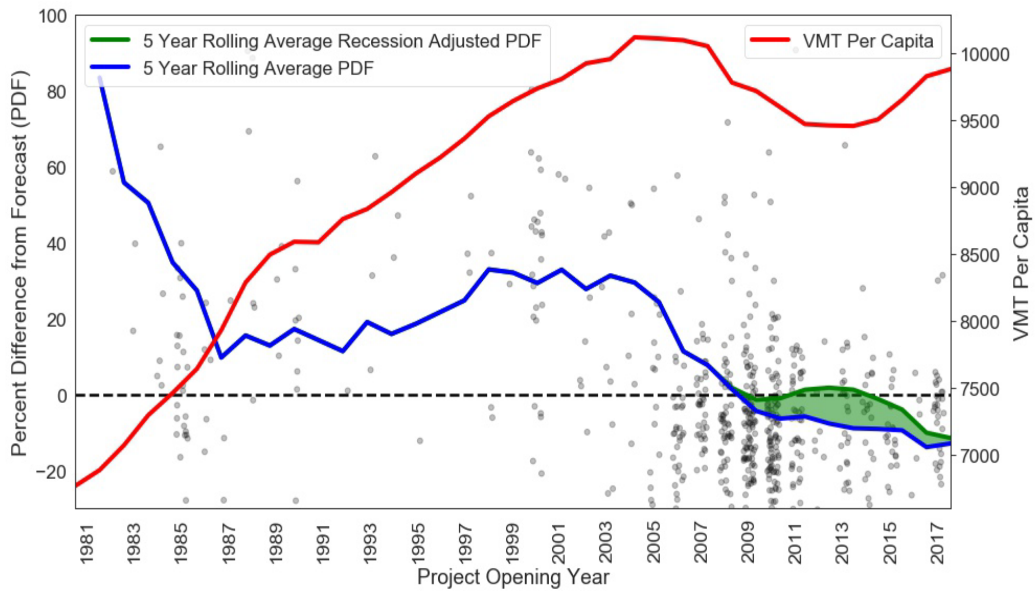

Figure 14: Trend in Percent Difference from Forecast, excluding resurfacing projects

3.6 Summary of Findings

In this research we used a large database to explore the accuracy of road traffic forecasts and document the distribution of counted versus forecast traffic volumes. The descriptive statistics provide insight into the factors affecting forecast accuracy and the changes in accuracy over time. Because we selected projects based on the availability of data, and they did not constitute a random or representative sample of all projects, selection bias may influence these findings. A large portion of the sample comes from one agency that recorded nearly all forecasts since the early 2000s, but inclusion of projects from other agencies is limited to those having sufficient documentation. In addition, several key variables such as forecast method and agency type, are missing for portion of the sample.

The missing data are not randomly distributed and instead relate to the practices of the agencies recording the data. Furthermore, 38% of the projects in our sample didn't record a definite opening year and we created a buffer based on the project type to get the post-opening traffic count. Errors in specifying the project opening affect conclusions about induced demand and can influence forecast accuracy. Despite these limitations, it is appropriate to conclude:

- Observed traffic was 6% lower than forecast on average, but this difference is due to lower traffic following the Great Recession. If not for higher unemployment rates from 2008 through 2014, the traffic would have been 1% higher than forecast on average. This result was not consistent through time, however, as we note next.
- The mean absolute difference between measured traffic volumes and forecasts was 17%. In addition, 90% of opening-year traffic volumes were in the range of -38% to +37% of the forecast volumes. This spread of outcomes persists after adjusting for the shift due to the Great Recession, suggesting that there are reasons for inaccuracy beyond this unforeseen event. These values highlight that we should not consider traffic forecasts to be point estimate, but a range of possible outcomes.
- The average deviation changed direction: observed traffic on projects opening before 2003 was higher than forecast but starting in 2003 it was lower. This change is due in part to the effect of the Great Recession, but a notable shift in the average deviation remains even after adjusting for the effect of the economic downturn. Evolving forecasting methods, a different mix of projects, or exogenous trends could explain this shift. We observed this shift even when limiting the analysis to capacity expansion projects, suggesting that changing project types did not fully

explain the change. The data showed a possible relationship to aggregate VMT trends. When VMT per capita was uniformly growing from the 1980s through early 2000s, observed traffic was on average higher than forecast. In the 2000s, however, VMT per capita leveled off, declined, then again increased. During this period observed traffic was on average lower than forecast. Evidence suggests that economic and fuel price changes determine much of the VMT change (Bastian, Börjesson, and Eliasson 2016). Those same factors may also explain changing traffic forecast accuracy. Future research should aim to untangle these relationships.

- Traffic forecasts became more accurate over time. In addition to the changes in average deviation noted above, projects opening more recently had a narrower spread of outcomes. Better data and improved forecasting methods may lead to this improvement, or it may relate to broader socioeconomic and project type trends noted above.
- Traffic forecasts were more accurate for higher volume roads and higher functional classes. The counted volumes on collector and arterial roads were more likely to be lower than the forecasts and percent deviation from forecasts had a greater spread than those on freeways. These challenges may be due to limitations of zone size and network detail, as well as less opportunity for offsetting inaccuracies on smaller facilities.
- Traffic forecasts were less accurate as the time span lengthens. Forecasts depend on exogenous projections which are more uncertain further into the future. They also depend on estimated relationships between travel behavior and those

exogenous factors that may evolve over time. Put simply: it is easier to predict tomorrow's traffic than it is to predict traffic 10 years into the future.

- Travel models produced more accurate forecasts than traffic count trends. Travel models are sensitive to the underlying determinants of traffic growth, including land-use changes and road network changes, so they were more accurate than traffic count trends.
- Some 95% of project forecasts meet the "half-lane rule". Considering the level of service in each segment, the inaccuracy in forecast would not have affected about 95% of the projects to warrant additional or fewer lanes. A total of 84% of project forecasts fell within the maximum desirable deviation suggested by NCHRP report 765 (Horowitz et al. 2014). These deviations were unlikely to affect a project decision about the number of lanes on a highway.

The descriptive analysis presented here provides insight into the degree of confidence that planners and policy makers can expect from traffic forecasts. While traffic forecasts have improved, substantial deviation between counts and forecasts remains, and the data reveal several factors related to accuracy. Among these are economic conditions, and we found evidence of a major unforeseen event—the Great Recession—causing a systematic shift in accuracy. In the wake of a much different disruption due to COVID-19, our results should open a discussion on communicating uncertainty in forecasting. It is reasonable to expect that there may be some major disruptive event within the scope of our next long-range forecasts. Moreover, such events are not the only factors contributing to forecast inaccuracy as a substantial spread of percent difference from forecasts remains after adjusting for the recession. Factors like forecast methodology, forecast horizon and project type affect the accuracy as demonstrated in this paper, along with other unknown or unquantified factors. Instead of dismissing forecasts as inherently subject to error, we recommend that agencies make forecasts more useful and more believable by acknowledging uncertainty as an element of all forecasting. Forecasts should not be a singular outcome, but a range of possible outcomes. Planners can combine uncertainty windows with decision intervals to determine whether a forecast deviation would change a project decision (Anam, Miller, and Amanin 2020).

NCHRP 934 provides instructions for accessing and contributing to this repository and offers advice about establishing a systematic process of data collection and evaluation. Additional systematically collected data will enable future research to identify sources of inaccuracy, compare the accuracy of different types of travel models, and guide the development of more accurate forecasting methods. More accurate traffic forecasts and greater understanding of factors that influence accuracy will contribute to more efficient allocation of resources and build public confidence in the agencies that produce those forecasts.

Chapter 4 CAUSES OF DECLINE IN PUBLIC TRANSIT RIDERSHIP IN THE UNITED STATES

The methodology proposed in this research to quantify uncertainty in forecasts rely on past accuracy data. however, past returns do not guarantee future performance, and this data-driven approach may be limiting if the future looks discontinuous from the past. For example, the effect of self-driving vehicles may pose a risk to forecasts made for 2040, and outcomes for projects that have already opened cannot clarify that risk. If there was a systemic change in the way people use public transportation, the change needs to be addressed in the uncertainty analysis. One such challenge in transit ridership forecast is the decline in ridership across the United States from 2012 to 2018 despite widespread investment in transit service. While these trends are remarkably consistent across US cities, transit ridership in other countries has increased in the last several years, with the few countries experiencing ridership losses also suffering from poor economic conditions or substantial demographic changes. The US has experience a strong economy, stable demographics and improved transit service over this period, making these ridership losses surprising. In this chapter, I present the results of our investigations into the causes of the recent transit ridership decline. We show that expanded transit service and land-use changes increased ridership 4.7% on bus and 10.7% on rail. However, losses due to other factors exceed these gains. Ride-hailing is the biggest contributor to transit ridership decline over this period, reducing bus ridership by 10%. Ride-hailing's effect on rail varies by metropolitan area size: it has little effect on rail ridership in the largest metropolitan areas but decreases rail ridership 10% in mid-sized metropolitan areas. Lower gas prices and higher fares contribute to lower transit ridership, as do higher incomes, more teleworking and higher car ownership.

This work was funded by the Transportation Research Board through Transit Cooperative Research Program (TCRP) Project A-43. The final project report includes a summary of these findings, several case studies and recommendations for practice (Watkins et al. 2021). This paper extends that work by placing the results in the context of the academic literature, providing a more detailed description of the data and methods, and summarizing the findings for an audience beyond transit practitioners. The content of this chapter has been adapted from:

Erhardt, G.D., **Hoque, J.M.**, Goyal, V., Berrebi, S., Brakewood, C., Watkins, K.E., (in-press), "*Why has public transit ridership declined in the United States*?", Transportation Research Part A: Policy and Practice.

4.1 Introduction

Transit ridership has declined sharply in the wake of the COVID 19 pandemic as cities have gone into lockdown to stop the spread since March 2020. But even before this unexpected change due to public health concerns, mode level transit ridership in the United States had been on a downward trend from 2014 to 2018 by a varying degree. While the total ridership has increased by a meagre 0.3% in 2019 compared to 2018, bus and light rail ridership show no sign of picking up with decrease of 1.04% and 4.5% respectively (APTA 2020).

What caused this decline, however, are not as easily discernible. The factors that could explain ridership trends in the past, such as service supply, population, and employment level, are not fitting the current situation. Overall vehicle revenue miles of the transit agencies have rebounded to their 2010-level by 2015 after the drastic service cuts following the recession and have kept growing ever since. Meanwhile, urban population and employment rates have risen substantially in the same period. At the same time, new trends in technology, travel behavior, and transportation policy have emerged. Especially the advent of Transportation Networking Companies (TNCs) like Uber and Lyft, delivery services like Grubhub, DoorDash and Amazon and teleworking may have had a significant impact on transit ridership.

Although there is a growing body of research on these factors, we still lack a comprehensive understanding of the extent to which various factors impact transit ridership. Many of the strategies transit agencies are using to mitigate or reverse trends are not well understood from a ridership impact perspective. This study captures the factors responsible for the pre-pandemic decline in transit ridership across the metro areas in the US through a longitudinal study of mode level ridership from 2012 to 2018. The model results establish the sensitivity of transit ridership to changes in the descriptive variables (service miles, fares, population, presence of TNCs and shared mobility etc.) covering the period of recent decline. We then conduct a series of sensitivity tests of transit ridership against these variables to ascertain how much each factor contributes to the change in ridership. This high-level analysis ensures that the trends we are capturing are broadly applicable across the nation.

4.2 Literature Review

Our literature review identifies two primary categories of factors that affect transit ridership: factors that are controlled by the transit agencies and otherwise. Moreover, the factors themselves are either traditional or emerging in the recent years as travel behavior changes. Combined, we get four overarching categories, presented in Figure 15.

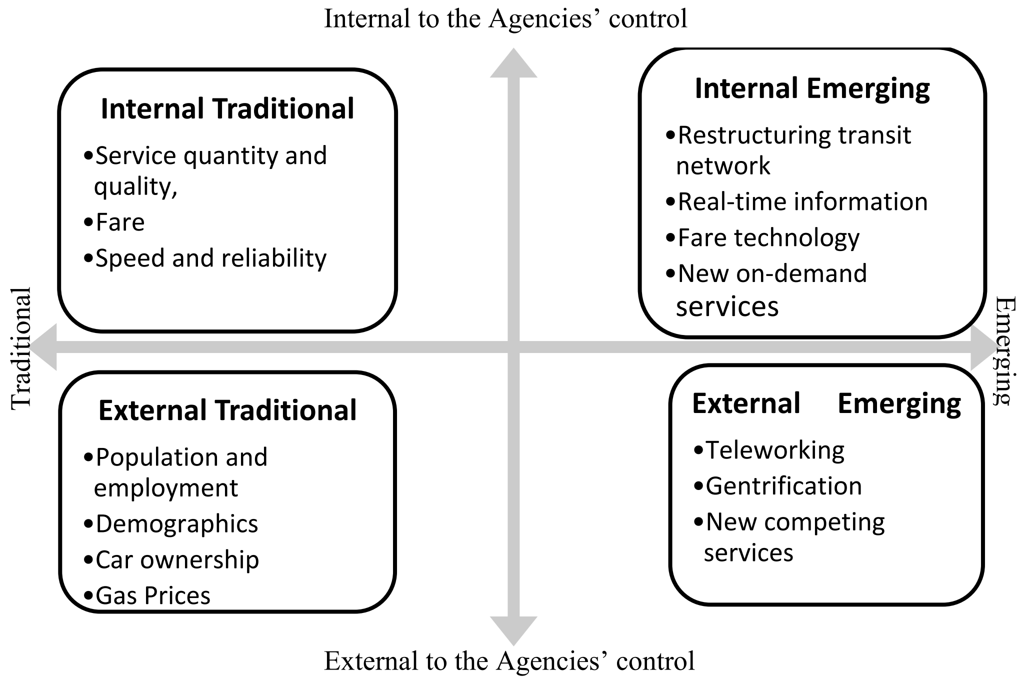

Figure 15: Factors Affecting Transit Ridership

The three primary areas under a transit agency's control that have traditionally impacted ridership are service quantity, fares, and service reliability. There is a consensus in literature that service levels, measured as Vehicle Revenue Hours (VRH) or Vehicle Revenue Miles (VRM), are the primordial factor affecting transit ridership (Kyte, Stoner, and Cryer 1988; Liu 1993; Gomez-Ibanez 1996; Kohn 2000; Evans IV 2004; Dill 2013; Boisjoly et al. 2018). Ridership is found to be modestly affected by frequency at the route segment level between 2012 and 2018 (Berrebi et al. In Review) and fare (Taylor et al. 2009; Chen, Varley, and Chen 2011; Mahmoud and Pickup 2019). A one percent change in either of these factors result in less than one percent change in ridership. Service reliability, measured as on-time performance, positively affects transit ridership as found by studies in Los Angeles (Chakrabarti and Giuliano 2015) and in Massachussetts (Thistle and Zimmer 2019).

Socio-economic and demographic trends and gas price, on the other hand, are factors outside the agencies' control that affect transit ridership. Transit ridership is positively correlated with employment level, despite it generating more commuting trips and private vehicle purchase (Hendrickson 1986; Z. Liu 1993; Gomez-Ibanez 1996; Taylor et al. 2009; Stanley 1998). The effect of gas prices has been found to vary based on their magnitude and mode (Nowak and Savage 2013), urban form (Maley and Weinberger 2009; Lee and Lee 2013), and timeframe (Yanmaz-Tuzel and Ozbay 2010). The impact however is relatively little on mode shift behavior, though they may cause some change in travel behavior in the short term when gas prices spike. Population makeup, particularly the share of Millenials (born 1980-2000), who exhibit a propensity for shared mobility (Grimsrud and El-Geneidy 2013; 2014) that can be in competition with transit (Alemi et al. 2018) may play a significant role as well (National Academies of Sciences, Engineering, and Medicine 2018). This effect may also be the manifestation of their tendency to move to auto-oriented suburbs as they settle family households.

These traditional internal and external factors identified in previous studies, however, do not fully explain the recent changes in transit ridership. Ridership have declined despite a 5% increase in bus service between 2012 and 2017. ("The National Transit Database (NTD) | FTA" n.d.). Furthermore, (Watkins et al. 2020) found that the relative change between VRM and Unlinked Passenger Trips (UPT) between 2012 and

2016 was loosely correlated at the metropolitan area level. Meanwhile, urban population in the United States is at its highest point in recorded history (US Census Bureau 2012) and urban core areas have grown in population every year since 2006 (Frey 2018). While (Driscoll et al. 2018) pointed out that transit-oriented regions are losing population and car-oriented regions are gaining them, (Watkins et al. 2020) finds that population change and ridership change were entirely uncorrelated for bus and somewhat correlated for rail, especially during this period of decline. On top of that, unemployment rate in 2017 in the United States were at their lowest level since the recession in 2009, suggesting there are emerging factors both within and outside the transit agencies' control that are influencing transit ridership.

Some of such factors identified in recent studies are changes in the network, availability of real-time transit information and new fare technology. Bus network redesigns increase ridership, but largely through increases in service and decreases in coverage (A. Schmitt 2017). The provision of real-time transit information was found to correlate with an increase in ridership (Tang and Thakuriah 2012; Brakewood, Macfarlane, and Watkins 2015b). While the impact of smartphone-based fare payment system on ridership remains unquantified, it is expected that the convenience it brings to a tech-savvy populace should have a positive influence on ridership.

The recent changes to how people travel also affect transit ridership. Telework, flex work schedules, and online shopping are becoming more prevalent and impacting the demand for travel or the times we do it. Another change in aggregate travel behavior is influenced by the advent of new mobility options, be it ride-hailing or Transportation Network Companies (TNCs), bike-shares or dockless scooters. Some see these new services as competitors that simply pinch riders from the transit system, while others believe that offering as many mobility options as possible enables individuals to choose a car-free or car-lite life. Longitudinal studies conducted at the transit agency or metropolitan area-level have come to diverging conclusions. Several studies using data up to 2015 have found that the entry date of Uber was had either a positive relationship with transit ridership or no statistically significant relationship (Hall, Palsson, and Price 2018; Boisjoly et al. 2018). Using a similar methodology but more recent data, (Graehler, Mucci, and Erhardt 2019) found that ride-hailing was correlated with a decline in transit ridership. While the evidence thus far seems to point towards ride-hailing as a potential cause of nationwide ridership decline, this relationship is still not well understood.

On the other hand, bike and scooter sharing systems can potentially enable first-mile/last-mile connectivity in suburbs and substitute transit in dense urban areas (D. Buck et al. 2013; Fuller et al. 2013; Martin and Shaheen 2011; Shaheen et al. 2014). Bike sharing system was associated with decreased bus ridership in New York (Campbell and Brakewood 2017) and increased Metrorail ridership in peripheral neighborhoods in Washington D.C. (Ma and Knaap 2019). The effect of dockless scooters is unquantified however, due to the recentness of the phenomenon. (Clewlow 2019) reported that in 11 major U.S. cities, 70% of the surveyed see electric scooters as a complement to public transit. (NACTO 2018) reported that in 2018, 25% of scooter trips are connections to transit. These results indicate that scooters may be enabling more ridership than they substitute. These findings, however, are only based on surveys and may be impacted by selection bias.

Another potential contributing factor to the decreasing transit ridership is the economic displacement of low-income earners, the primary transit user, from dense urban centers to the suburbs (Florida 2017). A study from (Berrebi and Watkins 2020b) find that a drop in the proportion of minority residents in Miami explains part of the ridership decline but not in Portland, Minneapolis, and Atlanta.

Although there is a growing body of research on these factors, we lack a comprehensive understanding of their contributions to recent transit ridership losses. Many of the existing studies focus on measuring the effect of a single factor, treating the others as control variables (Mahmoud and Pickup 2019; Chakrabarti and Giuliano 2015; Nowak and Savage 2013; Maley and Weinberger 2009; Lee and Lee 2013; Yanmaz-Tuzel and Ozbay 2010; Grimsrud and El-Geneidy 2013; Driscoll et al. 2018; Hall, Palsson, and Price 2018; D. Buck et al. 2013; Fuller et al. 2013; Martin and Shaheen 2011; Campbell and Brakewood 2017; Ma and Knaap 2019; Brakewood, Macfarlane, and Watkins 2015a; G. D. Erhardt et al. 2021). Other studies examine trends in a single location, which is valuable, but the findings may or may not apply elsewhere (Kyte, Stoner, and Cryer 1988; Mahmoud and Pickup 2019; Chakrabarti and Giuliano 2015; Thistle and Zimmer 2019; Nowak and Savage 2013; Maley and Weinberger 2009; Yanmaz-Tuzel and Ozbay 2010; Grimsrud and El-Geneidy 2013; 2014; D. Buck et al. 2013; Fuller et al. 2013; Campbell and Brakewood 2017; Ma and Knaap 2019). Several of the more comprehensive studies of the determinants of transit ridership pre-date the recent period of steep decline (Z. Liu 1993; Gomez-Ibanez 1996; Kohn 2000; Dill et al. 2013; Boisjoly et al. 2018; Taylor et al. 2009; Chen, Varley, and Chen 2011).

In this study, we quantify the effect of a broad set of factors on transit ridership, considering how some either offset or compound the effects of others and how their contributions may differ by location. Our results provide the most comprehensive understanding to-date of the contributors to the pre-COVID decline in transit ridership in the United States.

4.3 Clusters of Metropolitan Statistical Areas

While investigating the potential causes of ridership losses, we need to keep in mind the different environment where transit agencies operate. The travel behavior, and by extension transit ridership, in a dense metropolitan like Washington, D.C. will not be the same as that in a sparsely populated urban area like Lexington, Kentucky. This context affects not only the contributors to changing ridership, but also which strategies may be effective at offsetting ridership declines. Several studies have proposed "peer groups" of agencies using different metrices— by geographic region, demographic and operational characteristics (Perk et al. 2004), by population in the Metropolitan Statistical Areas (MSAs) (Brown and Neog 2012), and by metropolitan area population, percent of population living in a dense area, percent of zero vehicle households, and transit-agency operating expenses (Ederer et al. 2019b). The analyses point to non-uniformity in ridership changes across mode and groups or clusters. The American Public Transportation Association (APTA) also proposed a set of clusters based on the operating expense of the transit agency as well as the influence of external factors that favor transit ridership and competitiveness of transit compared to other modes.

The high operating expenses group (greater than $300 million annually) includes 19 MSAs with populations between 2 million and 13 million—such as Atlanta, Chicago, Philadelphia, and Houston—each with both bus and rail services.

The medium operating expenses group (between $30 million and $300 million annually) includes 64 MSAs ranging with populations between 200,000 and 4.6 million, such as Bakersfield, California; Denver, Colorado; Indianapolis, Indiana; and New Haven, Connecticut. All MSAs with mid operating expenses have bus service, and 12 of them also have rail service.

The low operating expenses group (below $30 million annually) includes 126 MSAs with populations ranging from 80,000 to 1 million—such as Athens, Georgia; Bridgetown, New Jersey; Morristown, Tennessee; and Yuma, Arizona—each with only bus service.

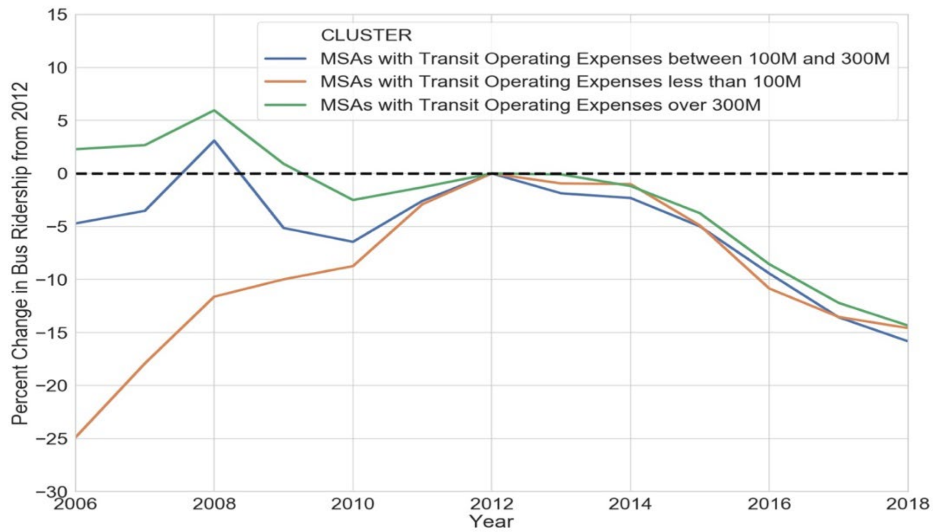

Figure 16: Percent Changes in Bus Ridership from 2012 by MSA Cluster

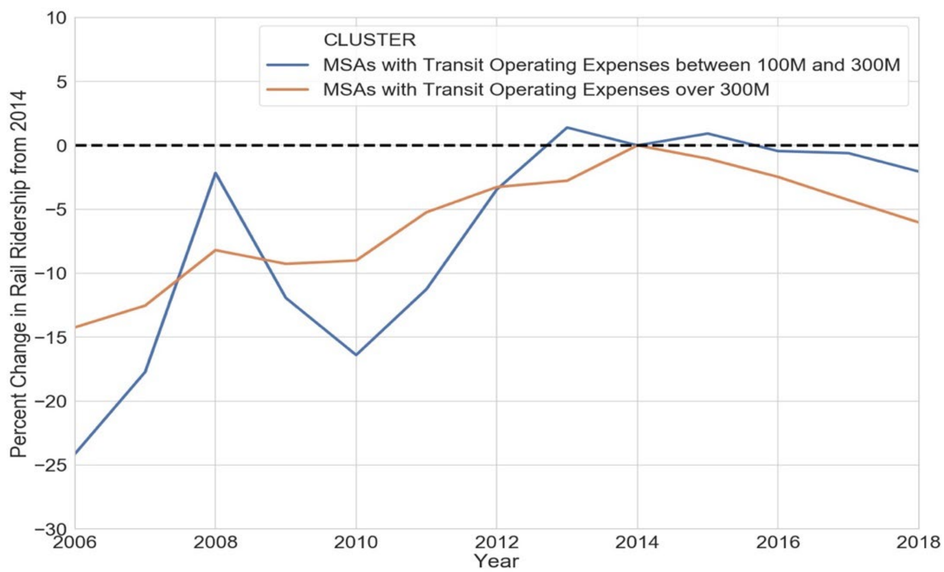

Figure 17: Percent Changes in Rail Ridership from 2014 by MSA Cluster

4.4 Data

The National Transit Database (NTD) reports time series data of transit profiles and summaries at an agency level, reported separately by mode. For each operator, we aggregated all types of bus (local bus, express bus, etc.) into a single bus mode and all types of rail (light rail, heavy rail, commuter rail, etc.) into a single rail mode. We exclude demand-responsive transit and all other modes. Often, multiple transit operators serve the same metropolitan area. We summed the unlinked passenger trips, vehicle revenue miles (VRM), vehicle revenue hours (VRH) and fare revenue for all operators within an MSA. Then we calculated average fare as the annual fare revenue divided by unlinked passenger trips. We resolved these differences in the boundaries of urbanized areas and MSAs by replicating Ederer et al's methodology (Ederer et al. 2019b). The resulting data file includes one record for each combination of MSA, year and transit mode (bus or rail). These combinations serve as the unit of all further analysis.

Individual transit agencies are responsible for reporting these data to the Federal Transit Administration (FTA) and the reporting is not always consistent. We manually reviewed the NTD data to identify such cases. For example, smaller agencies may report to NTD in some years but not others, some services changed names or merged with other operators, and sometimes the fare revenue is zero in one year but non-zero in all years both before and after. We manually reviewed the data to identify potentially anomalous cases and compared against local news reports and agency announcements to determine whether a jump in the data might correspond to a real-world service or fare change. In a limited number of cases, we either excluded problematic records, or interpolated values from the year before and the year after. We documented those decisions with notes in the estimation data file provided as supplementary materials.

Population and employment characteristics at the metro area level were obtained from American Community Survey (ACS) 1-year estimates from 2002 to 2017 and Bureau of Labor Statistics (BLS), with 2018 data extrapolated from the three previous years. A different dimension of the socio-economic and demographic variables is densities at transit supportive areas. We measured that using the Longitudinal Employer-Household Dynamic (LEHD) Origin-Destination Employment Statistics (LODES) dataset. The transit supportive densities are defined as census tracts having a total population and employment greater than 10 per acre.

One issue in conducting the analysis for MSAs is that because MSAs are defined by whole counties, a portion of the outlying areas may not be served by transit and there is not a simple mechanism to define the service area for all transit operators in the US. If much of the growth is in the outlying portion of an MSA, we would expect that growth to have little effect on transit ridership. To account for how centralized or dispersed the growth is, we identified census tracts as having transit supportive density if the 2010 population plus employment was greater than 10 per acre. We selected this threshold by mapping several breakpoints and based on a visual inspection selected one that captured a contiguous developed area for several metropolitan areas. This threshold is slightly higher than the minimum residential density of 3 households per acre (about 7.5 people per acre) suggested by the Transit Capacity and Quality of Service Manual (TCQSM) (Kittelson & Associates et al. 2013). The lower threshold resulted in a more patchwork map. We also found it important to include employment to avoid central business districts from being excluded. We measured the population from the decennial census and the employment using the Longitudinal Employer-Household Dynamic (LEHD) Origin-Destination Employment Statistics (LODES) dataset. We then used annual data from the LODES Workplace Area Characteristics to calculate employment in those transit supportive tracts versus other tracts in the MSA, and we used annual data from the LODES Residential Area Characteristics as a proxy for the population in those transit supportive tracts. We also compiled measures of the percent of low-income workers and employees, poor households, and minority households in transit supportive density to test the gentrification hypothesis.

Investigating the effect of ride-hailing services on transit ridership is a challenging task because of the absence of city-level ride-hailing trips data. This fact is echoed in (Boisjoly et al. 2018; Manville, Taylor, and Blumenberg 2018). TNC trips per capita, extrapolated from a National Household Travel Survey in 2017, revenue reported in Uber and Lyft common offerings and (Schaller 2018) have also been tested. However, this measure may not be a reliable predictor of TNC use because of sampling bias in the travel survey, assumptions of linearity across metropolitan areas and clustering effect. A reliable substitute is the number of years since the first arrival of ride hailing services in a metro area (Boisjoly et al. 2018; Graehler, Mucci, and Erhardt 2019), since it is strongly correlated with the penetration and growth of such services.

Uber staff provided the date in which they started operations in each city, and we used the years since Uber's arrival as a proxy for the number of trips. Given that the number of ride-hail trips continues to increase after its initial entry, it is preferable to use a proxy variable that also increases rather than a binary flag for ride-hail's presence (Graehler, Mucci, and Erhardt 2019; G. D. Erhardt et al. 2021). We also estimated the total ride-hail trips in the US in each year from revenue and trip data provided in the Uber and Lyft Initial Public Offering (IPO) documents (US Securities and Exchange Commission 2019b; 2019a). We allocated the 2017 ride-hail trips to MSAs proportionally to the number of trips that report taking taxi, limo, Uber or Lyft according to the 2017 National Household Travel Survey (NHTS), or proportionally to population where the NHTS data were sparse. Then we scaled the 2017 MSA level estimates to the annual US total to estimate the total ride-hail trips in each MSA. In this allocation, we ensure that trips are only allocated to MSAs where ride-hail is available in that year.

We acknowledge that each of these measures is an imperfect proxy for ride-hail ridership data. A better measure for use in this study would be the total number of ride-hail-trips served in each MSA in each year. Privacy and commercial interests are frequently cited as arguments against data sharing, but in this case such data are so aggregate that they would raise no privacy concerns, and it is not clear what commercial value they hold. While New York and Chicago have obtained ride-hail data through regulation (Taylor and Wasserman 2021), data elsewhere are not available to this study. Therefore, we proceed with these proxy measures and discuss the limitations in interpreting our results later in the paper.

We attempted to compile data on changing congestion from the Texas Transportation Institute's Urban Mobility Report (Schrank et al., 2019), but the data are not available for all MSAs and years in our sample.

We identified the presence of bike share and e-scooters in each MSA using data from the Bureau of Transportation Statistics (BTS) (BTS n.d.). These data only go back to 2015, so we identified the start dates of bike share systems that start prior to 2015 using local news reports.

We identified the year and location of bus network restructures from TCRP Synthesis 140 (Byala et al. 2019) and coded a binary variable indicating the restructure. We measured service quality and reliability as Mean Distance Between Failures, using mechanical and/or system failure as reported by the NTD. We did not have access to more comprehensive reliability measures such as on-time performance and such measures are not consistent across transit agencies. Safety incidents in 2015 and 2016 on the Washington Metro led to line closures and major maintenance work in the following years, with disruptions lasting from late 2015 to early 2018 (Delgadillo 2020; Duggan, Aratani, and McCartney 2016). To capture the effect of these disruptions, we coded a variable for rail in Washington, DC with a value of, 0.5 in 2015 when the fires disrupted the system, 1 in 2016-2017 when the most extensive track closures took place, 0.5 in 2018 when track work continued, and 0 otherwise. The variables tested and their changes between 2012 and 2018 are listed in Table 9.

Table 9: Descriptive variables and their changes across MSA clusters

Variable	MSA Operating Expense Cluster	Mode	2012	2018	% Change
Ridership (Unlinked Passenger Trips, 000s)	High	Bus	133,740	114,547	-14%
		Rail	88,648	86,115	-3%
	Medium	Bus	15,019	12,649	-16%
		Rail	6,283	5,912	-6%
	Low	Bus	2,338	1,996	-15%
Service Supply (Vehicle Revenue Miles)	High	Bus	42,251	45,087	7%
		Rail	25,489	29,502	16%
	Medium	Bus	7,074	7,876	11%
		Rail	2,173	2,722	25%
	Low	Bus	1,306	1,434	10%
Service Supply (Vehicle Revenue Miles)	High	Bus	42,251	45,087	7%
		Rail	25,489	29,502	16%
	Medium	Bus	7,074	7,876	11%
		Rail	2,173	2,722	25%
	Low	Bus	1,306	1,434	10%
Total Population	High	Bus	4,942	5,266	7%

Variable	MSA Operating Expense Cluster	Mode	2012	2018	% Change
		Rail	4,942	5,266	7%
	Medium	Bus	1,202	1,290	7%
		Rail	1,648	1,780	8%
	Low	Bus	324	343	6%
Total In-state Population	High	Bus	42,251	45,087	7%
		Rail	25,489	29,502	16%
	Medium	Bus	7,074	7,876	11%
		Rail	2,173	2,722	25%
	Low	Bus	1,306	1,434	10%
Service Supply (Vehicle Revenue Miles)	High	Bus	42,251	45,087	7%
		Rail	25,489	29,502	16%
	Medium	Bus	7,074	7,876	11%
		Rail	2,173	2,722	25%
	Low	Bus	1,306	1,434	10%

4.5 Methods

The sensitivity of transit ridership to changes in the descriptive variables is established through a longitudinal analysis of mode level transit ridership. Such relations vary across the metro areas as well as over time and are estimated through a Panel Ordinary Least Squared (OLS) Model.

Transit ridership is in essence a demand-supply problem. The relative utility of transit compared to the other modes depends on the supply (frequency, density of stops, accessibility, proximity to attractions etc.) as well as the fare. This supply is in turn dependent on the ridership—the more people using the service, the more the agencies are prompted to increase their service. This endogeneity violates the basic assumption of regression. In addition, it is not possible to include every factor in the analyses: as described in the previous section, we didn't consider several variables in our dataset because of their unavailability. These omitted variables are also likely to interact with the other variables in the model, producing biased estimation. Assuming unobserved factors at each MSA that might simultaneously affect the ridership and the demographic variables do not change over time, we consider Fixed Effect in our model estimation. Fixed effect models avoid the unobserved heterogeneity and endogeneity biases by using each individual entity as their own control in time. Fixed effects models control for the effects of time-invariant variables with time-invariant effects. This is true whether the variable is explicitly measured or not. If y_{it} is the total ridership, or UPT, for Metropolitan Statistical Area *i* at year *t,* and x_{it} are the explanatory variables, the standard format of fixed-effect Panel OLS is:

$$y_{it} = \beta_0 + \beta_1 x_{it} + \sum_{j=1}^{n-1} d_j \alpha_j + \varepsilon_{it} \qquad Equation\ 1$$

Where d_j is a dummy variable equal to 1 for MSA *j* and 0 for the others. There are n-1 dummy variables, one for each MSA except the last one whose fixed-effect is merged with the constant term. α_j is the fixed effect for MSA *j*.

The model itself can take different forms. Since ridership is essentially a count data, it is skewed, and its variance increases with their mean. Skewed data can be transformed using the natural logarithm if they have constant variance to the mean. In our analysis we estimated a mixed log-log and log-linear model noting the non-linear relationship of ridership with k dependent variables as well as the skewness of the data.

$$\log\,(y_{it}) = \beta_0 + \beta_1 \log(x_{kt}) + \beta_2 x_{(i-k)t} + \textstyle\sum_{j=1}^{n-1} d_j \alpha_j + \varepsilon_{it} \qquad Equation\ 2$$

Taking the exponent,

$$y_{it} = e^{\beta_0} * x_{kt}{}^{\beta_1} * e^{\beta_2 x_{(i-k)t}} + u + \in \qquad Equation\ 3$$

The coefficients of the regression represent elasticity of ridership against the log transformed explanatory variables. For the non-transformed variables, each unit increase in X multiplies the expected value of Y by e^{β}. For example, a coefficient of -0.27 on log-transformed Average Fare means that for 1% increase in the fare, ridership decreases by 0.27%. For the non-transformed variables, each unit increase in X multiplies the expected value of Y by e^{β}, in other words expressing a percent change in Y with a unit change in X. For example, increasing the linear variable Percent of Zero-Vehicle Households with a coefficient of 0.01 means that with each unit change, ridership increases by 1%.

4.6 Model Estimation Results

The OLS regression results of the model specified by Equation 2 is presented in Table 10. The model estimates the log-transformed Unlinked Passenger Trips of 215 MSAs as a function of several explanatory variables. The shaded cells indicate coefficients that are statistically insignificant.

Table 10: Fixed-effects panel data model of the log of bus and rail ridership in each MSA

Dependent Variable	Transf.	Entities	R-squared
Unlinked Passenger Trips	Log	240	0.54
Description	**Transf.**	**Coeff.**	**t-statistics**
Service			

Vehicle Revenue Miles (Bus)	Log	0.449	14.66
Vehicle Revenue Miles (Rail)	Log	0.662	16.05
Major maintenance event		-0.133	-1.89
Network restructure		0.047	1.35
Fare			
Average Fare (in 2018$) (Bus)	Log	-0.579	-16.29
Average Fare (in 2018$) (Rail)	Log	-0.346	-4.3
Land Use			
Population + Employment	Log	0.218	2.78
Percent of total employees living and working in Transit Supportive Density in an MSA		0.399	1.39
Gas Price			
Average Gas Price (in 2018$)	Log	0.143	7.77
Household and Income Characteristics			
Median Per Capita Income (in 2018$)	Log	-0.071	-1.19
% of Households with 0 Vehicles		0.002	0.78
% Working at Home		-0.008	-2.86
New Competing Modes			
Effect of the Presence of TNCs on Bus Ridership			
At MSAs where transit operating expenses exceed 300M		-0.019	-4.71
At MSAs where transit operating expenses are less than 300M		-0.033	-12.66
Effect of the Presence of TNCs on Rail Ridership			
At MSAs where transit operating expenses exceed 300M		0.002	-0.46
At MSAs where transit operating expenses are between 30M to 300M		-0.023	-3.85
Presence of Bike Share		-0.011	-1.51
Presence of Electric Scooters		-0.039	-3.28

The specific variables used in the analysis are described below and grouped into six broad categories. In all cases, when discussing change, we refer to net changes, assuming that all other factors remain constant.

4.6.1 Discussion of Variables

The variables included in the final model can broadly be categorized in several groups based on their effects. Service refers to the factors internal to the transit agency: how many vehicle miles they are operating their transit modes, whether or not there were any major service disruptions because of maintenance or any network restructuring. The land use category includes population and employment and their concentration in the metro area. These factors are external to the transit agency's control but are strong determinants of transit service supply. Transportation Network Companies (TNCs), Bike and E-Scooter sharing services have revolutionized the transportation infrastructure and in previous research they have been found to compete with public transit. We have grouped them together in the New Competing Modes category.

4.6.1.1 Service

Vehicle revenue miles (VRM) of service is a strong determinant of transit ridership. The results indicate that each percent increase in bus VRM increases bus ridership by 0.45% and each 1% increase in rail ridership increases rail ridership by 0.66%. Rail ridership may be more elastic to changes in VRM because it tends to attract more choice transit riders than bus.

Bus network restructures are associated with 4.7% higher bus ridership, but the effect is not statistically significant. In recent years several transit operators have restructured their bus network, changing routes and the service allocation, to better serve their passengers. The operators that made these changes saw, on average, a 4.7% bus ridership increase over and above the effect of any VRM increases. However, not enough agencies have completed such a restructure to make the result statistically significant.

Major line closures for maintenance work can have an important effect on rail ridership. Safety incidents in 2015 and 2016 on the Washington Metro led to line closures and major maintenance work in the following years, with disruptions lasting from late 2015 to early 2018. We found that rail ridership in the Washington MSA was 13% lower in the affected years (with half the effect in 2015 and 2018) than would otherwise be expected. This effect was marginally significant. We tested a more comprehensive measure of reliability based on the mean distance between failures (MDBF), but the reporting of failures to the NTD is inconsistent and we could not detect a meaningful effect.

We tested or considered several other measures of transit service. We found that the average transit speed was negatively correlated with transit ridership, probably because vehicles can travel faster if they do not have to stop to pick-up and drop-off passengers. We could not have a widely available measure of on-time performance, nor did we have a comprehensive measure of where the service is allocated within a region.

4.6.1.2 Fare

Higher fares lead to lower transit ridership. Increasing average bus fare by 1% decreases bus ridership by 0.57% and increasing average rail fare by 1% decreases rail ridership by 0.35%. The average fare is calculated by taking the total inflation-adjusted fare revenue (inflation adjusted) earned in a year by the agency per unlinked passenger trip. The different elasticities for bus versus rail fare may reflect different income mixes of the passengers. We could not test specific fare or pass programs at the system level.

4.6.1.3 Land Use

Each 1% increase in population plus employment is associated with 0.22% more transit ridership. These effects are correlated with each other and could not be estimated separately, but when taken together the effect is positive and significant.

Higher density leads to more transit ridership. For each percentage point increase (such as from 10% to 11%) in population plus employment living in the transit supportive areas, transit ridership is 0.4% higher.

Working at a national level, we could not compile data on the location and size of transit-oriented developments, or other more detailed data on the allocation of land use within transit supportive areas.

4.6.1.4 Gas Price

Each percent increase in gas price accounts for 0.14% increase in transit ridership. Increase in gas price induces people to rely more on public transit rather than privately owned auto. We adjust the measure for inflation.

4.6.1.5 Household & Income Characteristics

With higher per capita income, people are less likely to ride transit. We have tested several variables to establish the relation between income and transit ridership. Although mean and median values of household level income display expected correlation, we chose per capita median income in 2018 dollars because of better fit of the model. Each 1% increase in the median per capita income results in 0.07% decrease in transit ridership.

Higher shares of 0-vehicle households in an MSA, have a small positive effect on transit ridership. We know that people from households without a car constitute an important market of transit riders. However, the share of 0-vehicle households has been relatively stable in recent years, so the results show that it explains little about the *change* in transit ridership over this period. Our results show that an increase from 10% of households owning 0 vehicles to 11% of households owning zero vehicles would result in 0.2% more transit ridership, but this effect is not statistically significant.

For each additional percent of workers telecommuting, transit ridership decreases by 0.76%. This result is based on the journey-to-work mode shares reported in the American Community Survey. This result is particularly interesting going forward considering the large percent of population working from home during the COVID-19 pandemic.

We tested the percent of the population living in poverty, the percent of the population born in a different country, and the percent of the population in different age groups and did not find significant effects. We also tested the distribution of poverty as measured by the percentage of poor households living in areas with transit supportive density but did not find a significant result.

4.6.1.6 New Competing Modes

TNCs negatively affect both bus and rail ridership. The effect is noticeably large and statistically significant in the MSAs with transit operating expenses between 30 to 300 million. In large metro areas with significant transit service already present, effect of TNCs is low.

Presence of TNCs affect the transit ridership negatively, more with each year in the market. The market penetration of TNCs increase every year after arrival until they reach an equilibrium/saturation point. We tested this ramp-up effect on bus and rail separately and found it has a stronger effect on bus ridership than on rail ridership. The model results show that TNCs have little and statistically insignificant effect on transit ridership in the High Operating Expense cluster. Bus ridership decreases by 0.7% while rail ridership decreases by 0.2% for each unit increase in TNC trips per capita. The smaller change in rail ridership can be attributed to the first and last mile connectivity of TNCs while they replace some bus trips because of their shorter coverage. The effect of ride-hailing service is more pronounced in the second cluster of MSAs with operating expense less than 300 million. Every year the ride hailing services result in 3.3% and 2.2% decrease in bus and rail ridership in these MSAs respectively.

Existence of bike sharing system (dockless and otherwise) positively affect bus and rail ridership, albeit with small statistical significance. On the other hand, e-scooters negatively affect both. A point to note here is that e-scooters are very recent addition to the transportation troposphere— the earliest of them in our dataset have been introduced in 2018. So, their effect may not be noticeable for some time into the future. Combined effect of bikes and scooters didn't produce any significant result.

Removing TNCs from the model to see the effect of bike and scooter share produces higher negative values for these variables. This suggests that since both are happening at the same time, it is difficult to separate the effect of each. When we remove bike share, the coefficients on the TNC variables don't change significantly, suggesting the TNC variables are more stable or more important than bike share variables.

4.6.1.7 Other Factors Tested

Rail competition and 1 year rail ramp up period has been considered to evaluate the effect of a competing and a newly introduced rail mode respectively. Hypothetically, the existence of a rail mode will attract some customers from bus. A new rail mode will take a while to achieve the necessary level of attraction. These variables were removed from the model because of the complexity they introduced without any noticeable benefits.

We have also tested the effect of immigrant population (percent of the population not born in the USA), as well as the effect of network restructure and maintenance. Their effects were insignificant to include in our final model. Effect of the age of the population and the poverty level in the MSA were found to be insignificant as well.

4.7 Contribution to Ridership Decline

We applied the sensitivities calculated above to calculate the total contribution of each of these factors to the change in transit ridership between 2012 and 2018. The coefficients for each variable in our estimation represent percent change of transit ridership for each percent or unit change in the explanatory variable. We multiplied these coefficients by the observed change in each factor to calculate that factor's effect on transit ridership. As we did so, we calculated the change in each explanatory variable from the previous year and its contribution to ridership change, following the approach used previously (G. D. Erhardt et al. 2021). Because the dependent variable is log-transformed, the exponential of this term gives a ratio that can be used to factor the ridership from the previous time period, holding all other terms constant:

$$\frac{y_{it}}{y_{it-1}} = e^{\beta_l * (x_{lit} - x_{lit-1})}$$

We label a specific factor's contribution to ridership change as the factor affecting change, or FAC. We calculated the FAC of factor k as:

$$FAC_{kit} = y_{it-1}\left(e^{\beta_k * (x_{kit} - x_{kit-1})} - 1\right)$$

After calculating the FAC separately for each variable, we label any remaining change as the unexplained change, FAC_u:

$$FAC_{uit} = y_{it} - y_{it-1} - \sum_k FAC_{kit}$$

which is similar to a residual change. To obtain the values reported in Table 2, we summed across entities and time periods:

$$FAC_k = \sum_{i \in M} \sum_{t \in T} FAC_{kit}$$

where T is the set of years from 2012 through 2018 and M is the set of entities specific to bus or rail. To calculate the charts in Figure 2 and Figure 3, we tabulate the cumulative FAC for a set of factors and add or subtract those from the observed ridership in each year. Whereas we used a log-log model to reduce the skew of the data and estimate direct elasticities, we reported the FAC results in units of ridership for a more intuitive interpretation. Therefore, when we sum across MSAs, those MSAs with more transit ridership have a greater influence on the totals, which explains why the total effect of ride-hailing on rail ridership is positive, even though ride-hailing has a significant negative effect on rail ridership in medium sized MSAs. While competing factors may offset each other, we used this approach to calculate the effect of each. Applying this approach did not capture 100% of the observed ridership change, and we labeled any remaining difference between the modeled and observed ridership as "unexplained change". We applied these calculations separately for each MSA and transit mode (bus vs rail) then summed across MSAs, excluding New York, for the results reported in this paper.

Table 11 shows the change in each factor and its contribution to bus and rail ridership change between 2012 and 2018.

Table 11 Contributions to bus and rail ridership change from 2012 to 2018

Description	Bus Ridership		Rail Ridership	
	Change in Average Value	Effect on Ridership	Change in Average Value	Effect on Ridership
Service				
Vehicle Revenue Miles	5.5%	3.1%***	12.5%	10.3%***
Network Restructure	0.02	0.1%*		
Major Maintenance Event			0.05	-1.0%
Subtotal		3.3%		*9.3%*
Fare				
Average Fare (2018$)	5.7%	-0.6%***	10.7%	-2.6%***
Subtotal		-0.6%		*-2.6%*
Land Use				
Population + Employment	6.6%	1.5%***	6.0%	1.4%***
% of Pop+Emp in Transit Supportive Density	-0.8%	-0.1%	-0.8%	-0.007%
Subtotal		1.4%		*1.4%*
Gas Price				
Average Gas Price (2018$)	-28.2%	-3.6%***	-28.5%	-3.7%***
Subtotal		-3.6%		*-3.7%*
Household & Income Characteristics				
Median Per Capita Income (2018$)	10.3%	-0.7%	10.5%	-0.8%
% of Households with 0 Vehicles	-8.9%	-0.2%	-9.8%	-0.2%
% Working at Home	29.5%	-0.8%***	28.1%	-0.9%***
Subtotal		-1.7%		*-1.9%*

Description	Bus Ridership		Rail Ridership	
	Change in Average Value	Effect on Ridership	Change in Average Value	Effect on Ridership
New Competing Modes				
Years Since Ride-Hail Start	4.27	-10.6%***	5.04	0.8%
Bike Share	0.69	-0.8%	0.57	-0.7%
Electic Scooters	0.34	-1.6%***	0.6	-2.4%***
Subtotal		-13.0%		*-2.3%*
Total Modeled Ridership		-14.1%		0.2%
Total Observed Ridership		-14.7%		-3.0%
Unexplained Change		-0.7%		-3.2%

Asterisks indicate that the estimated coefficient is statistically significant for a 90%(*), 95%(**) or 99%(***) confidence interval

We find that two sets of factors pushed to increase transit ridership over this period:

- **More service.** Transit operators provided service in the form of added vehicle revenue miles (VRM). Following deep bus service cuts in the aftermath of the 2008 recession, bus VRM increased 5% leading to 3% higher bus ridership, which further increased due to several bus network restructures. These years continue a three-decade period of investment in expanded rail service with 12% more rail VRM between 2012 and 2018, resulting in rail ridership increases of 10%. These gains were offset slightly by major rail maintenance disruptions in Washington, DC. These service additions varied substantially by location depending on the service provisions of the local operators.

- **Land use.** Land use also affects transit ridership, both in terms of total population and employment growth, and how centralized that growth is. Population and employment in these MSAs grew an average of about 6% between 2012 and 2018, with that growth slightly less centralized. The combined effect of land use changes were bus and rail ridership increases of 1.4%. It makes sense that these effects were modest, because while land use is an important driver of transit ridership, changes tend to occur over a long time frame.

The causes of transit ridership decline between 2012 and 2018 came from a combination of four main sources. Together, these sources more than offset the factors above that pushed ridership up over this period. They include:

- **Income and household characteristics.** Higher incomes, higher car ownership, and an increase in the percent of people working at home contributed to bus and rail ridership declines of about 2%.

- **Higher fares.** Fare increases were operator-specific, so the effect varies by location, but fares were on average higher in 2018 than in 2012 after adjusting for inflation. Average bus fares increase 6% and average rail fares increased 11% leading to 0.6% lower bus ridership and 2.6% lower rail ridership.

- **Lower gas prices.** Average inflation-adjusted gas prices decreased by more than a quarter over this period, leading to between 3% and 4% lower bus and rail ridership.

- **New modes compete with transit.** Three new modes emerged in cities over this period that compete directly with bus: ride-hailing, bike share and e-scooters. The analysis shows that the effects of bike share systems and e-scooters were much smaller compared to ride-hailing services. Ride-hailing itself contributed to 10% lower bus ridership, with the combined effect of all modes leading to 13% lower bus ridership. For rail, the effect of ride-hail varied by MSA size. For MSAs in the high operating expenses group, ride-hailing's introduction increased rail ridership by an insignificant amount, but in mid-sized cities, ride-hailing reduced rail ridership by 10% on average. Because the larger MSAs have much higher rail ridership, the overall effect when we combine across all MSAs is slightly positive. The combined effect of all three new modes led to 2% less rail ridership, although the bike share effect is statistically insignificant and data on electric scooters is limited to a single year.

In Figure 18 and Figure 19 we applied the model to each year from 2012 through 2018 and plotted the effect of each factor on ridership for bus and rail, respectively. We observe that expanded service was the largest contributor to ridership gains for both bus and rail. Lower gas prices starting in 2014 led to ridership losses. For bus (Figure 18), we observe that new competing modes were the largest contributor to ridership loss, and that we would expect ridership to be roughly flat if not for this new competition.

The model suggests that when we consider all of these factors together, we would expect bus ridership to have declined by 14.1% and rail ridership to have increased slightly by 0.2%. In comparison, observed ridership decreased by 14.7% and 3% for bus and rail respectively, leading to -0.7% unexplained changes in bus ridership and 3.2% in rail ridership. It is not surprising that this model does not fully capture the changes to rail ridership because there are fewer MSAs with rail, and rail systems in the US are diverse—they include heavy rail systems many decades old, newly constructed light rail systems, commuter rail and more. The smaller number of observations makes it more difficult to capture some of the dynamics that may affect rail differently. However, it is important to note that considering the large expansion of rail service over this period, we should expect a corresponding ridership increase. The fact that rail ridership declines despite its expansion is quite striking, and the model does capture most of this difference.

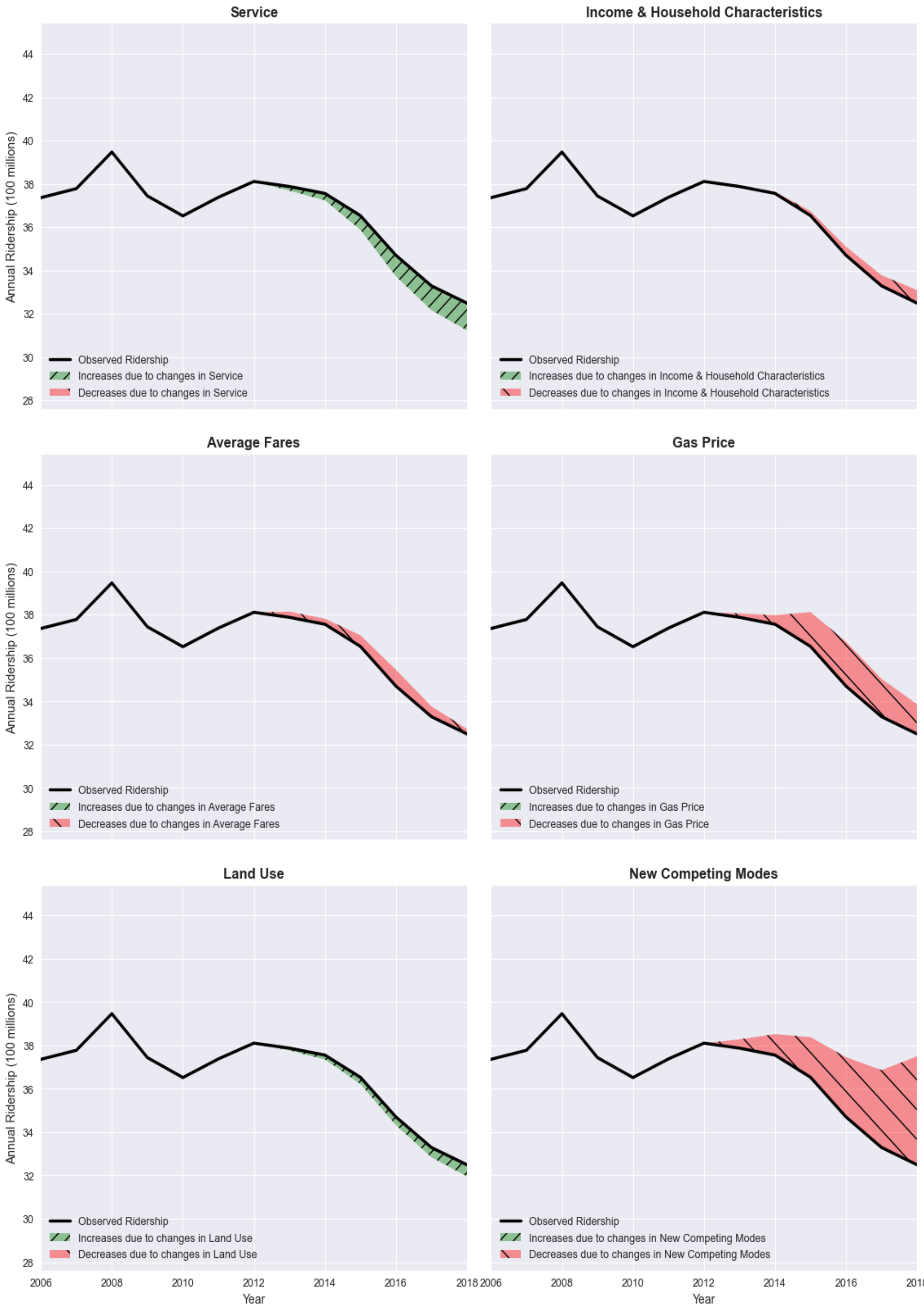

Figure 18 Contributions to bus ridership change relative to 2012.

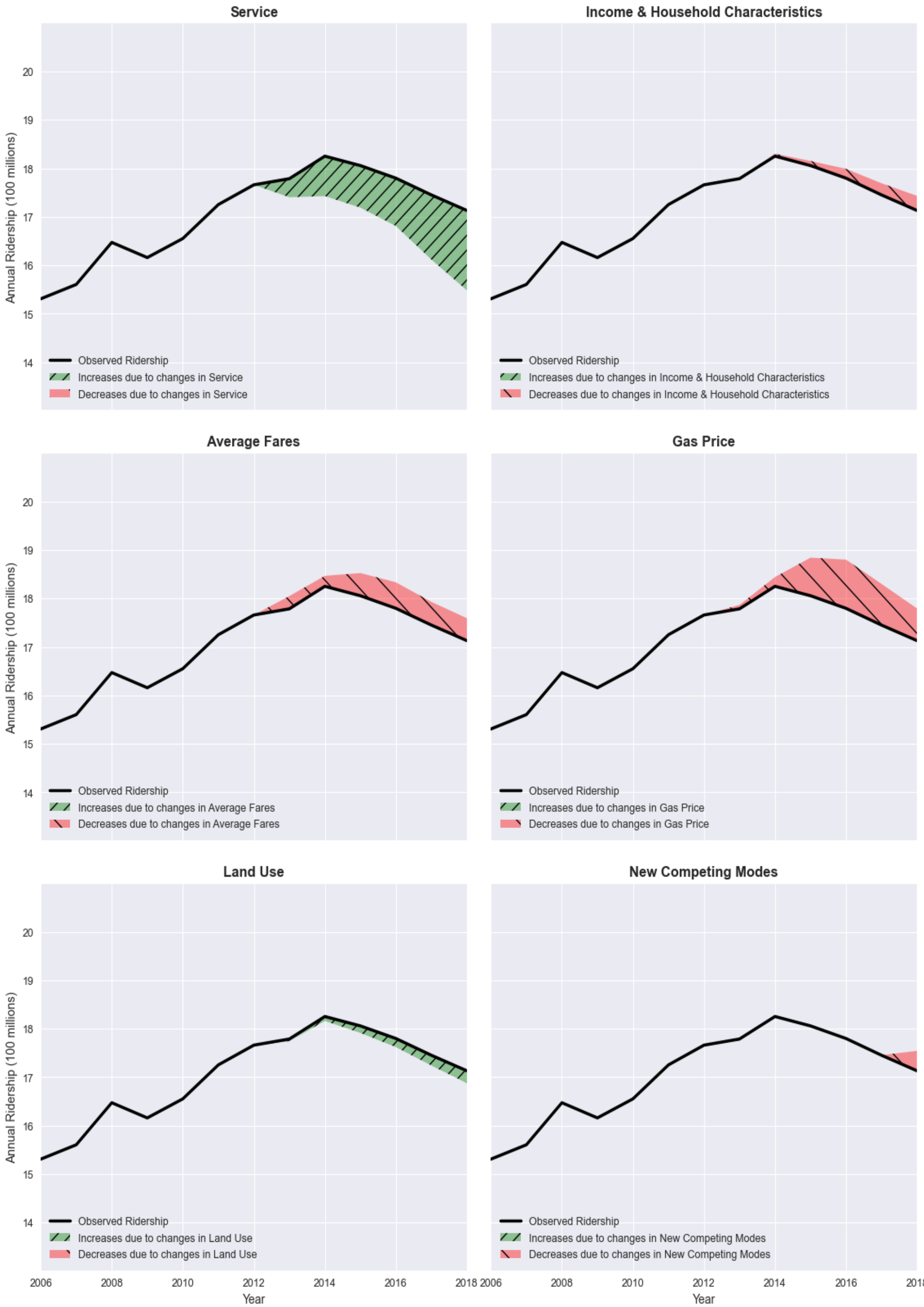

Figure 19 Contributions to rail ridership change relative to 2012.

4.8 Limitations

An important limitation of our study is that we use years since market entry in each MSA as a proxy for ride-hail ridership. The risk of using this proxy is that it is potentially capturing some other unrelated change. We explored the implications of this risk in the preceding section and find that though there is uncertainty in the magnitude of the ride-hail effect, ride-hailing has a consistent negative correlation with transit ridership for three out of four market segments. The exception to this rule is for rail in the high operating expenses group, where the ride-hailing coefficient has a positive coefficient in some tests but a negative and significant coefficient when we consider the estimated number of trips rather than the years since entry. Nonetheless, there are enough assumptions built into our estimates of ride-hail trips that we prefer the simpler proxy. If data on the total number of ride-hail trips by MSA becomes available, we recommend this study be repeated to take advantage of those data.

We recommend caution interpreting the results beyond the study period. When the COVID-19 pandemic hit in 2020, both transit ridership and ride-hail use dropped (L. Liu, Miller, and Scheff 2020; Loa et al. 2022), so it would not make sense to extrapolate the years since market entry variable through this period. By summer 2021, anecdotes suggested that ride-hail prices were much higher than before the pandemic (Paul 2021), which could limit the number of ride-hail trips if prices remain high.

There are several additional variables that may be important but that we could not effectively capture. These include road congestion and transit on-time performance. It is possible that as congestion increases over this period, the buses become less reliable, which might explain in part why rail ridership, which is more frequently on dedicated right-of-way, declines less than bus. Transit on-time performance data is not consistently available across regions and not even consistently measured by transit agencies. Given that e-scooters are only available in the final year of our analysis, we are not confident in the estimated effect. In a separate, more detailed, analysis of the effect of e-scooters on bus ridership in Louisville, we found no measurable effect on local bus ridership and a possible complementary effect on express bus ridership (Ziedan et al. 2021).

As noted in the methods section, transit ridership and the supply of transit service are endogenous: ridership is higher when agencies provide more service, and agencies are motivated to provide more service in areas where ridership is high. We mitigated this problem by using a fixed-effects model that effectively estimates the coefficients based on the change in the value of each term from the year before. Therefore, we must consider whether the change in ridership and the change in service are endogenous, such as if operators add service because ridership is growing or cut service because ridership is decreasing. However, the motivation for this paper is the opposite—operators added service while ridership decreased—so the risk of this result being driven primarily by endogeneity appears low. Nonetheless, is worth comparing our estimated service elasticities to others reported in the literature. Taylor et al., (2009) evaluated the determinants of transit ridership in 265 US urbanized areas and used two-stage simultaneous equation models to account for endogeneity between transit supply and ridership. They find that the change in headway with respect to changes in service frequency across all transit modes is about 0.5—between our estimated bus and rail service elasticities.

More broadly, there are limits to what can be measured at such an aggregate measure, so more detailed studies within cities, such as Berrebi and Watkins, (2020b), Erhardt et al., (2021) and Ziedan et al., (2021) complement these findings.

4.9 Summary of Findings

The decline in transit ridership during 2012 to 2018 despite investments into expanding and modifying the service and positive socio-economic trends has been baffling. Our analysis shows that while this infusion of funds and population growth have had a net positive effect on bus and rail ridership, they are offset by factors outside the transit agencies' control. The steep decline of about 14.7% in bus ridership is largely influenced by the presence of TNCs. As the market penetration of TNCs rose each year, they started replacing more and more bus trips contributing to about 10.6% of the decline. They have a net positive effect on rail ridership (0.2%) however, indicating first and last mile coverage enabled by TNCs. But TNCs contribute to about 10% decrease in rail ridership in areas where transit operating expenses are below 300 million dollars per year. Similarly, presence of bike sharing systems and e-scooters have a negative effect on ridership across modes and clusters.

While some factors identified in previous works remain unquantified because of lack of data, the results present transit operators and transportation planners insight into developing new strategies to respond to the declining ridership. The fundamental motivation for these strategies needs to base on ensuring equity and social justice: providing a travel option especially for those without other means of travel and providing a resource-efficient and climate-sensitive means of moving volumes of people. Basic transit service expansion (increased routes, frequencies, spans) could increase ridership simply by adding more service. In addition, including such expansion strategically may improve productivity (ridership per trip) of certain routes. Ride-hailing & car sharing partnerships could help transit agencies retain riders by using the best of these services in partnership with transit. Similarly, addressing first and last mile access to transit is critically important to retaining or improving ridership. Demand response services, flex route services, and micro-transit pilots can help serve the first and last mile.

Just as one of the traditional factors is the impact of increases in fares will cause ridership to decrease, fare discounts or reduction/elimination is a strategy agencies could pursue to increase ridership. Fare policy innovations can target specific segments of the populations through targeted fare discounts by time of day (e.g., weekend passes) or type of customer (e.g., social fares, off-peak senior fares, etc.). Although fare reduction or removal is not often used in the US, targeted fare discount initiatives are growing. Fare free zones have also been used in several agencies across the country.

Of course, the short and long-term effect of the COVID-19 pandemic need to be weighed in as well. The Great Recession of 2008 shows that economy takes a while to recover, and this may pose certain challenges and opportunities for the transit agencies. Considering transit as a social service, the transit agency can focus specifically on prioritizing essential workers and travelers with limited options, providing access to jobs and services. Even though they may be least likely to leave transit, these riders may be the most important to serve.

Chapter 5 THE ACCURACY OF TRANSIT RIDERSHIP FORECASTS

Accuracy of transit ridership forecasts have garnered attention over several decades. Historically they have been found to be optimistically biased, even more so than traffic forecasts as we have discussed in the previous chapter. However, they have been getting better over the years, with increased focus on getting the uncertainties inherent accounted for in the forecasts. The Federal Transit Administration (FTA), in this regard, has been at the forefront with a systematic forecast accuracy review program as part of the Capital Investment Grant program. In this chapter, I analyze the overall trend of transit ridership forecast accuracy across the years based on the largest database of transit ridership forecasts and contextualize it with the recent developments as described in the previous chapter. I find that transit ridership is about 24.6% lower than forecast on average with about 70% of the projects over-predicting ridership. Forecast accuracy varies by mode, area characteristics, length, time span and horizon. The accuracy has been getting better over the years, particularly after 2000 with the introduction of new analytical and evaluation tools. The steadily improving accuracy, however, is offset by the unexpected decline since 2012 as explored in Chapter 4. When we adjust ridership for the changes in metro area unemployment, auto-ownership, median income, gas prices and presence of Transportation Network Companies, the aggregate accuracy improves. Even so, there remains substantial deviation in the observed ridership from forecasts.

5.1 Introduction

The importance of public transit in urban transportation planning cannot be overstated, especially in terms of ensuring climate-sensitivity and social equity. Public transit has a much lower carbon-footprint per passenger and provides a viable mode of transport for people without access to a car. Investments in such infrastructure is informed by travel demand forecasting models that drive the benefit-cost analysis. Inaccuracy in these forecasts can therefore skew the cost estimates against a projection of benefits. For several decades, studies have investigated forecast accuracy in tolled and un-tolled road traffic and transit ridership. For transit infrastructure projects in the USA, observed ridership has typically been about 16% to 44% lower than forecast (Webber 1976; Pickrell 1990; Kain 1990; Button et al. 2010; Nicolaisen and Driscoll 2014; D. Schmitt 2016). Similar level of inaccuracy in the global context of transit ridership forecasts is reported in (Bent Flyvbjerg, Holm, and Buhl 2005; Nicolaisen 2012) as well. Recent studies, however, have found that accuracy in transit ridership forecasts are getting better over time in the USA with the advent of new and improved analytical tools and better scrutiny of the models themselves, particularly for projects part of the New Starts program (Lewis-Workman et al. 2003; 2007; Voulgaris 2019a). Project mode (Button et al. 2010; Voulgaris 2019a), construction time (Voulgaris 2019a), presence of an existing system (Button et al. 2010) and when the project was constructed (D. Schmitt 2016) have been found to be statistically significant in their effect on transit ridership forecast accuracy. From the context of travel demand forecasts in general, the accuracy has been found to be a function of project characteristics, exogeneous inputs and the model parameters and specifications themselves (Hugosson 2005; Hoque, Erhardt, Schmitt, Chen, Chaudhary, et al. 2021a; Zhao and

Kockelman 2002). Recent studies have found that ridership has been at a decline since 2012, even before the start of the pandemic induced lull (G. D. Erhardt et al. Submitted; Berrebi and Watkins 2020a). Several recent factors have contributed to this decline, namely the advent of shared mobility and ride-hail services as well as lower gas prices and higher income levels (G. D. Erhardt et al. Submitted). Since such changes are unexpected, it stands to reason that the accuracy of ridership forecasts particularly during these years are also affected by these factors.

The meta-analysis of demand forecast accuracy in (Nicolaisen and Driscoll 2014) notes that accuracy is much lower for transit ridership forecasts compared to traffic. Several explanations are possible:

- There could be a methodological difference such that bus and rail are more difficult to predict for technical reasons having to do with them being lower-share alternatives, the difficulty of estimating good values-of-time, or the challenges associated with identifying transit markets or transit users.
- It may be that rail and toll road projects only get built when the forecasts show strong demand, whereas un-tolled road projects tend to get funded regardless. This could lead to optimism bias in the forecasts, as suggested by (Bent Flyvbjerg 2007a) or it could lead to self-selection bias, as suggested by Eliasson and Fosgerau (2013), where projects with forecasts that happen to be too low don't get built, and therefore don't end up in the sample.
- It could also be that the long-term trends over the past 40 years associated with growing auto ownership, the entry of women into the workforce, and

> high levels of suburbanization combined to create a future that was not anticipated at the time the forecasts were made but is systematically biased to push people towards using roads and away from transit.

Moreover, the decline in transit ridership in the US from 2012 to 2018 may have an impact on their forecasts as well. There is an absence of rigorous statistical analysis to identify and quantify the impact of different factors affecting the inaccuracy. With well over 200 large-scale transit projects constructed since the 1970s, there is surprisingly little publicly available data on demand forecasts from transit projects beyond those that receive large federal grants from FTA's Capital Investment Grant program. In this study, we employ the largest known database of transit ridership forecasts in the United States. It comprises a meaningful sample of all constructed large-scale public transit projects. The database contains information on several project and forecast characteristics in addition to actual ridership. In this study, we will be focusing on the accuracy of forecasts by mode and over the years across these projects and forecast characteristics in addition to the factors identified to be affecting the recent trend in ridership.

5.2 Data

According to Transport Politic, approximately 283 unique projects have been constructed between 1974-2019 in the United States ("The Transport Politic - Transit Explorer 2021" n.d.). We based our analysis of transit ridership accuracy and uncertainty on a database of 164 large-scale transit projects across the United States. The database is compiled through personal efforts by Mr. Dave Schmitt and is currently the largest known database of this kind. The projects include downtown people movers, Bus Rapid Transit

(BRT), Light Rail Transit (LRT), Heavy Rail Transit, and commuter rail projects. Information contained in the database include, but are not limited to, project and forecast characteristics like length, location, mode, service area and travel time characteristics, observed ridership where available and exogeneous forecasts like cost estimates, population, and employment projections etc. In addition, we have also made use of the set of projects included in (Voulgaris 2019a) to fill out missing fields and add more projects in the dataset.

A limitation of the database is the high degree of missing data on key variables. Because of the absence of standardized reporting of project and forecast information, such data are often not recorded in the project documents released to the public. The projects span five decades, from the 1970s to the 2010s. Projects built since 2000 comprise over 70% of the database.

Unfortunately, there is no standardized reporting of key inputs and forecasts in post-opening analysis or news articles. A further challenge is that the accuracy detail of the inputs varies greatly. For example, projects analyzed by FTA through their Predicted versus Actual or Before and After Studies are more likely to have explicit, numerical information about the accuracy of the inputs. For other projects, the accuracy is described qualitatively. In these cases, the accuracy level is determined qualitatively by reading the text. Inputs not reported in documentation, other reports or news articles are not reported in these tables. Consequently, even when data is available some input types are poorly represented.

The socio-demographic data have been collected at the Metropolitan Statistical Area (MSA) level from the American Community Survey (ACS) data, and the Bureau of

Labor Statistics (BLS) data. However, the MSA delineation have changed over the years and data before 2005 and after 2019 were not available at the time of analysis. In such case, we used linear interpolation from the decennial census data to fill the blank fields. Such interpolation introduces additional bias in the analysis as these are different from the data used in the models. However, they do present the opportunity to evaluate the changing accuracy as the demographics shift over the years.

Table 12: Data Description

Field	Definition	Availability
Forecast Ridership	Forecast Ridership in average weekday for a project.	
Actual Ridership	Observed Ridership in average weekday for a project.	
Project Development Phase	Defined as the planning phase in which the forecast was made. Planning/environmental, engineering/design and funding decision phase.	
Year Forecast Produced	The year the forecast was generated.	
Forecast Year	The future year for which the forecast was generated.	
Year of Observation	The year that actual ridership was observed. Many projects have multiple observed ridership values. Actual ridership from the year closest to the forecast year is used.	
Ramp Up	The number of years after project opening that the observation is taken.	
Mode	Primary mode of the transit system. Can be one of Bus, Light Rail, Commuter Rail, Downtown People Mover, Streetcar/Trolley and Urban Heavy or Light Rail.	
Number of stops	The number of stops added/served by the project.	
First mode	Whether the project introduces first of its kind in the system.	
Length	Length of the transit system.	
Servicing Central Business District	Whether the project services the central business district.	
Service Level	The project’s assumed frequency. Actual Value as a percentage of assumed value.	

Travel Time	Time to travel from end to end. Actual Value as a percentage of assumed value.
Fare	Project fare per unlinked passenger trips.
Supporting transit systems	Existing transit systems in the service area.
Gas Price	Gas price in the year forecast was produced or the year of observation in the Metropolitan Statistical Area (MSA), adjusted to 2019 dollars. Obtained from the Energy Information Administration (EIA). In USD (inflation adjusted).
Per Capita Median Income	Median Income of individuals or households adjusted to 2019 dollars. Obtained from the American Community Survey 1-year estimation for the Metropolitan Statistical Area the project is located in. In USD (inflation adjusted).
MSA population and population characteristics	Total Population in the Metro Area. Other variables tested: racial mix, percent of immigrant population, percent of population born in and out of the state of residence, poverty status, and age distribution. Source: ACS 1-year estimation
MSA employment and employment characteristics	Total unemployment rate in the Metro Area. Source: Bureau of Labor Statistics.
MSA household and household characteristics	Household characteristics (% of 0-vehicle households, household median income). Source: ACS 1-year estimation
Area Transit Characteristics	Defined by the yearly transit operating expense of the MSA. The MSAs are divided into three broad categories: large, multi-modal MSAs which spend more than $300 million a year on transit operation, medium sized MSAs spending between $100 to $300 million yearly and the smaller MSA with less than $100 million transit operation expenses.

5.3 Method

For measures of accuracy in transit ridership forecasts to set the context of uncertainty analysis, we define the accuracy as Percent Difference from Forecast (PDF) as in Equation 2:

$$PDF = \frac{Observed\ Ridership - Forecast\ Ridership}{Forecast\ Ridership} * 100\%$$

Negative values on the metric indicate that the observed ridership was lower than the forecast, and positive values suggest the opposite.

For apples-to-apples comparison, it is imperative that the forecasts and the observation to be compared are on the same year. However, transit ridership usually undergoes a "ramp-up period" where people learn about the new service and adjust travel behavior accordingly (Chang et al. 2010). The Before and After Studies done by the FTA for their CIG program recognizes this effect and considers the ramp-up period of usage maturity in their evaluation. FTA compares the actual ridership measured via on-board surveys conducted two years after project opening to opening year forecast ridership (Federal Transit Administration 2020). Moreover, (Shinn and Voulgaris 2019) presents statistically significant evidence of ridership ramp-up affecting forecast performance. It also shows that the effect is realized by the second year after project opening and therefore, considering it as the observation year may be appropriate for forecast performance measurement. In light of this evidence, we have considered a maximum of two years of ramp-up, i.e., two years after project has opened, for analyzing the accuracy. This reduced the sample size to 136 projects from our initial sample of 164.

As we know from the project development life cycle, forecasts are made at different phases in the planning process. In our database, we have several ridership forecasts made at different project development phase. For consistency, we considered the forecast at the latest available stage of the cycle. Most often, this is the funding decision phase, as the forecast for the design phase are typically optimistically biased to avoid under-designing. For apples-to-apples comparison, the forecast and observed ridership needs to be in the same year as well. In this study, we take a different approach to that of FTA's and instead carry on with our approach of comparing forecasts and observation at the same year. In case the observation is at a later year, we extrapolated the forecast using to be at the same year as observation. This extrapolation, however, introduces a bias as it does not consider the ramp up effect of transit ridership. After applying such selection criteria, we based our analysis on a reduced sample of 125 projects, all of which has an observation and a forecast ridership in the same year.

5.4 Transit Ridership Forecast Accuracy

Overall, our sample has a mean PDF of -24.6%, signifying that transit ridership forecasts are higher than the observed ridership on average. Almost 80% of the projects have had ridership less than the forecast value. The Mean Absolute PDF (MAPDF) is 40.2% which signifies the average deviation of observed ridership from the forecasts. About 90% of the projects in the sample have seen ridership deviating between -81.6% and 45% from forecast, represented by the 5^{th} and 95^{th} percentile values. This spread, along with the standard deviation of 46.4 indicate high variability in the forecast performance. Moreover, actual ridership has rarely exceeded the forecast for the projects even considering several years of ramp-up effect. Even so, the actual ridership increases with each additional year after project opening, diminishing the deviation from forecasts. This suggests that ridership forecasts are highly uncertain and optimistically biased to the point that the forecast demand does not realize several years after opening. However, we need to consider the effect of other explanatory variables to come to a robust conclusion regarding this observation. Figure 20 presents the overall distribution of the percent difference from forecast.

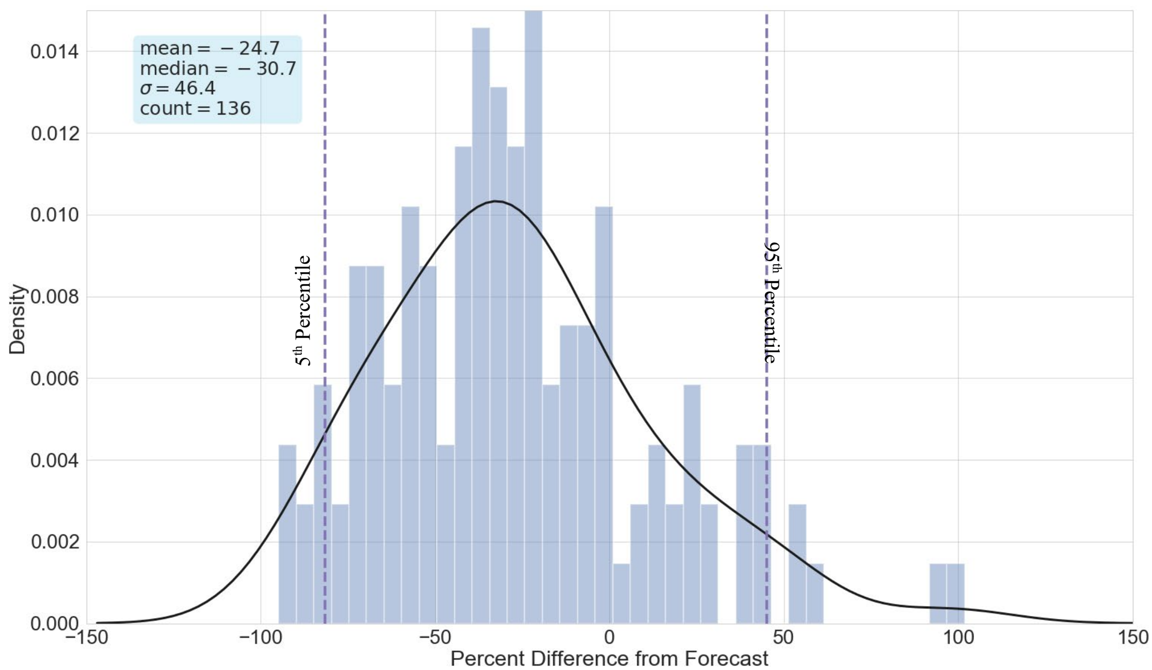

Figure 20 Percent difference of observed ridership from forecast

The statistical measures of transit ridership forecast accuracy across different categories are presented in Table 13. In the rest of the section, we present some of the key observations.

Table 13: Percent Difference from Ridership Forecast by Category

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
Overall Distribution							
Overall	136	41.2%	-24.7%	-30.7%	-81.6%	45%	63.3%
By Ridership							
Less than 5,000	44	40.1	-13.60	-23.88	-74.31	57.58	65.9
5,000 to 10,000	20	40.7	-8.07	-13.01	-76.05	30.75	53.4
10,000 to 15,000	20	38.5	-34.69	-30.11	-85.71	-1.46	42.1
15,000 to 20,000	12	35.3	-23.42	-23.39	-70.79	28.15	49.5
20,000 to 25,000	7	34.1	-31.80	-38.10	-59.45	1.64	30.5
25,000 to 30,000	5	41.9	-10.53	-12.24	-72.37	46.61	59.5
30,000 to 50,000	12	34.0	-30.40	-32.54	-51.59	9.64	30.6

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
More than 50,000	12	53.9	-53.92	-60.96	-69.71	-22.21	23.7
Ramp Up							
0 to 2 Years of Ramp Up	136	41.2%	-24.7%	-30.7%	-81.6%	45%	63.3%
3 Years of Ramp Up	50	38.8%	-29.5%	-29.9%	-80.2%	34.1%	57.2%
4 Years of Ramp Up	38	35.6%	-20%	-21.4%	-76.2%	52.4%	64.3%
5 Years of Ramp Up	26	31.1	-20.5%	-19.8%	-73.9%	27.2%	50.6%
5+ Years of Ramp Up	22	39.7%	-19.3%	-25.5%	-76.6%	66.1%	71.4%
Mode							
Bus	2	71.2	-71.2	-71.2	-80.0	-62.5	8.8
Bus Rapid Transit	32	40.0	-7.6	-19.3	-65.8	67.0	66.4
Commuter Rail	22	36.1	-34.2	-36.3	-74.2	6.5	40.4
Downtown People Mover	4	82.9	-82.9	-82.1	-92.2	-74.7	8.8
Streetcar/Trolley Rail	18	43.6	-14.8	-25.0	-87.1	58.4	72.7
Urban Heavy Rail	15	48.5	-48.5	-54.8	-71.6	-18.7	26.4
Urban Light Rail	43	35.9	-20.6	-26.5	-73.3	44.8	59.0
Project Length							
less than 5 miles	45	44.7	-27.0	-34.1	-85.6	45.2	65.4
between 5 to 10 miles	30	39.2	-25.4	-26.7	-77.6	42.0	59.8
between 10 to 20 miles	33	41.0	-13.8	-26.5	-68.3	42.0	55.1
greater than 20 miles	28	38.1	-32.9	-36.4	-74.3	11.3	42.8
Year Forecast Produced							
Before 2000	39	46.3	-41.9	-49.2	-81.4	11.1	46.3

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
2000 to 2008	30	30.7	-19.4	-23.3	-66.3	40.6	53.5
2008 to 2012	22	44.6	-24.6	-33.3	-82.7	43.8	63.2
2012 to 2015	7	27.0	-14.4	-21.4	-37.3	27.0	32.2
After 2015	5	46.2	-25.6	-32.4	-69.3	35.6	52.4
Year of Observation							
Before 2000	20	48.1	-44.2	-50.1	-81.1	1.8	41.5
2000 to 2008	36	50.4	-29.4	-39.4	-85.5	29.8	57.6
2008 to 2012	21	24.8	-14.7	-20.0	-43.9	40.0	42.0
2012 to 2015	22	34.1	-18.1	-15.3	-74.3	43.0	58.7
After 2015	37	42.1	-19.1	-28.6	-84.9	60.6	72.7
Time Span							
0 to 1 year	5	36.9	-16.2	-28.6	-42.1	37.0	39.6
2 years	4	41.0	-18.9	-22.0	-69.2	35.6	52.4
3 years	6	33.2	-33.2	-30.2	-62.8	-12.7	25.0
4 years	8	30.6	-17.1	-24.5	-53.0	30.7	41.8
5 years	13	35.2	-4.7	-17.8	-51.6	69.5	60.5
More than 5 years	67	43.0	-36.3	-37.2	-84.7	25.9	55.3
Project Jurisdiction by CBSA Transit Operating Expense							
Greater than $300m	73	40.1	-33.8	-34.8	-79.6	23.7	51.7
Less than $300m	55	37.6	-15.3	-20	-84	49.4	66.7
Service Area Characteristics							
In CBD	87	37.6	-20.9	-24.4	-82.5	49.7	66.1
Not In CBD	49	47.7	-31.4	-37.2	-80.1	31.0	55.6
First Mode							
First Mode	65	44.4	-17.9	-27.7	-81.4	57.2	69.3
Not First Mode	71	38.3	-30.8	-32.1	-80.9	24.3	52.6
CIG Project							
CIG Project	79	41.5	-32.9	-34.9	-81.2	39.1	60.2
Not CIG Project	57	40.8	-13.2	-21.5	-79.7	60.6	70.2

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
New Line							
New Line	30	34.2	-29.2	-28.0	-83.5	16.9	50.2
Not New Line	106	43.2	-23.4	-33.1	-80.5	50.0	65.2
Extension							
Extension	26	43.0	-32.3	-35.9	-77.7	39.9	58.8
Not Extension	110	40.8	-22.9	-28.3	-82.2	48.8	65.5

5.4.1 Mode

Forecasts for bus rapid transit, urban light rail and streetcar or trolley perform better on average than the others with a much lower mean PDF (-7.6%, -20.6% and -14.8% respectively). The spread of outcomes (represented by half the difference between the 5th and 95th percentile values) are at a similarly high level for these three modes, denoting significant variability in accuracy among the projects. The rail systems (commuter and urban heavy rail) perform better in this aspect, but they are highly optimistic (mean of -34% and -48.5% with 95% of the projects having PDF below 6.5% and -18.7% respectively). A reason for commuter and urban heavy rail having a large deviation from forecasts may be the scope of their service. These two modes typically serve longer routes with heavier traffic than light rail, streetcars, and people movers. The travel models used to forecast the ridership may not adequately account for the large network with high variability in demand in the analysis.

The average length for transit projects has been on the decline, i.e., recent transit projects are smaller in length and scope (Figure 21). Projects with a smaller length mean fewer stations and fewer ridership, in addition to less sensitivity to land-use and economic changes. Because of their length, streetcars and trolleys have therefore a smaller scope and more accurate forecasts. Commuter rails, on the other hand, typically serve longer distances and therefore have a much larger scope contributing to more degrees of freedom. It is possible that this length variable is interacting with other variables as well, since the crosstabulation in Figure 22 doesn't present any noticeable trend across different modes.

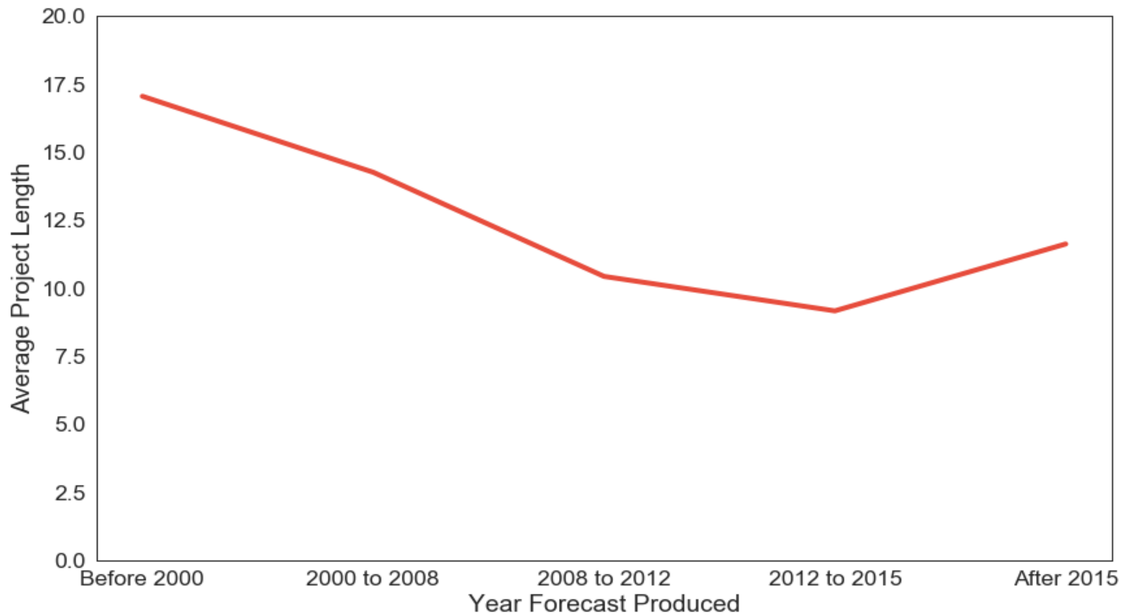

Figure 21: Average Project Length Over the Years

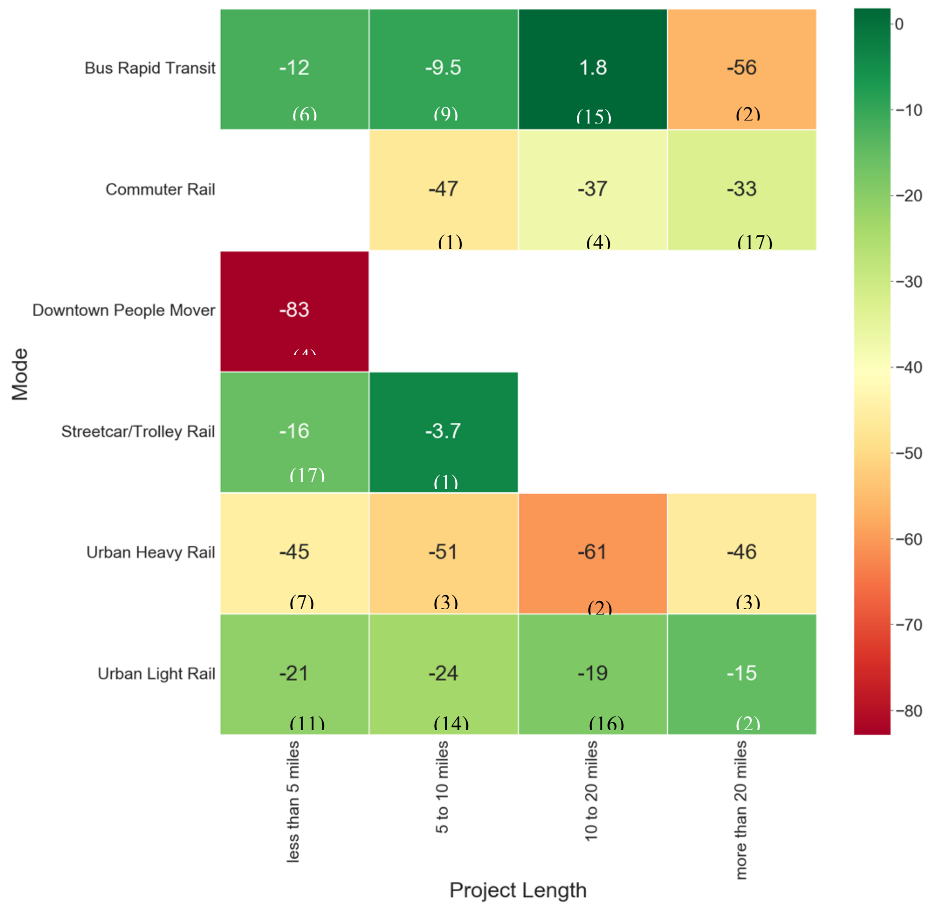

Figure 22: Forecast Performance by Mode and Length

However, the difference in forecast performance by mode can also be the effect of other external factors like year forecast was produced, transit ridership trend in the opening and observation year, project type and area transit characteristics etc.

5.4.2 Project Type

(Voulgaris 2019a) hypothesizes that forecasts on extensions and renovations of an existing system by adding new lines to it would be more accurate because of local experience with transit and the agency's familiarity with their forecasts. While the differences in mean PDF are not statistically significant, ridership on projects that do not create a new line on an existing network or extend a line have lower deviation from forecast. The spread of deviation remains between 50% to 65%. Ridership on such projects is closer to the forecasts suggesting ridership on existing network is more predictable. However, we cannot say the same for projects that add a new transit mode. Transit projects that are the first of their kind in a metro area perform better on average (mean PDF of -17.9% against -30.8%). We find this result counter-intuitive because we might expect it to be more difficult to forecast ridership on a newly introduced public transportation mode. Indeed, we find the average absolute deviation is higher for the first modal projects (44.4% against 38.3%), suggesting the lower average deviation may be an effect of positive and negative deviations cancelling each other out. Looking into individual projects that make up this category, we further notice that most of these projects are small and therefore a small change affects a large deviation from forecast. Again, the deviation from forecast can be the materialized effect improving forecasting methodology and transit trend over the years. Figure 23 presents the changes in average deviation over the years. We see that forecast performance for projects adding new line or extending service have improved over the years, although there is a noticeable change for projects opening from 2012 to 2015. As we have discussed in the previous chapter, transit ridership experienced a sudden and unexpected decline during these years which are presumably not accounted for in the forecasts.

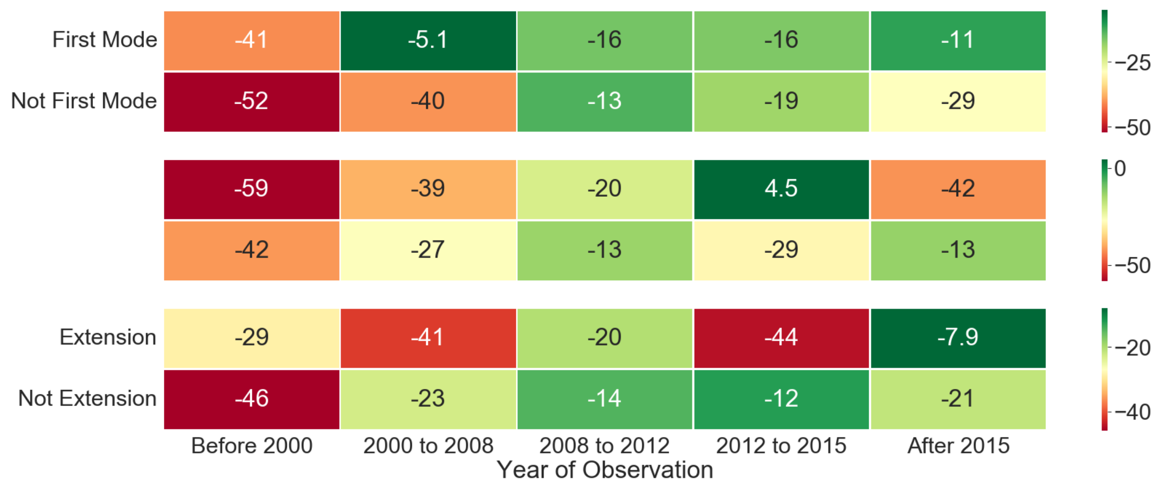

Figure 23: Ridership Forecast Performance by Project Type

5.4.3 Service Area

Projects serving the Central Business District have higher ridership. This means that ridership for these projects is closer to the forecasts (average deviation of -20.9% against -31.4%, with smaller absolute deviation). It is generally assumed that work travel patterns are easier to model than non-work travel because of the publicly available home-to-work records in the American Community Survey and Longitudinal Employee-Household Dynamic (LEHD) data. We found that transit projects that serve CBD areas have a narrower range of outcomes than the rest, indicating relative consistency. However, these projects still have a wide range of outcomes, suggesting that forecasting models still lack sufficient intricacy to address future demand.

Another important factor of transit ridership is the area's familiarity with transit systems. Project sponsors serving larger populations may have greater resources to devote to preparing rigorous forecasts. They may also answer to a wider variety of stakeholders, which could influence the incentives for promoting a particular project through optimistic forecasts. We tested this effect have considering the yearly operating expense of the transit agency which forms the basis of the cluster defined by the American Public Transportation Association (APTA). The high operating expenses group (greater than $300 million annually) includes 19 MSAs with populations between 2 million and 13 million—such as Atlanta, Chicago, Philadelphia, and Houston—each with both bus and rail services. The mid operating expenses group (between $30 million and $300 million annually) includes 64 MSAs ranging with populations between 200,000 and 4.6 million, such as Bakersfield, California; Denver, Colorado; Indianapolis, Indiana; and New Haven, Connecticut. All MSAs with mid operating expenses have bus service, and 12 of them also have rail service. The low operating expenses group (below $30 million annually) includes 126 MSAs with populations ranging from 80,000 to 1 million—such as Athens, Georgia; Bridgetown, New Jersey; Morristown, Tennessee; and Yuma, Arizona—each with only bus service.

Our results show that there is a statistically significant difference in the mean PDF for forecasts in metro areas with yearly transit operating expense on the two sides of $300 million. Ridership on larger metropolitan areas with significant transit presence deviates more from their forecasts (mean PDF of -33.8%) and about 88% of such projects had lower ridership than forecasts.

5.4.4 Service Area Characteristics

It is well understood from literature that service area socio-demographic characteristics have a bearing on transit ridership. Declining population and employment often have negative impacts, while that in aggregate income and auto-ownership result in a rise in transit ridership. Gas price also affects transit ridership by influencing riders to move away from driving cars. For optimistically biased forecasts, increased ridership would result in the average deviation from forecasts smaller. Our analysis show that MSAs that have seen growing unemployment rates, zero-vehicle households and gas prices from start to observation year has a smaller average PDF proving our hypothesis correct. This also indicates that the errors in exogenous forecast used in transit demand forecast have contributed to the accuracy, or the lack thereof. We could not glean any useful evidence from the population categories because of small sample size.

Table 14: Transit Ridership Forecast Accuracy by Service Area Characteristics

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
Metro Area Population Growth from Start to Observation Year							
Declining	3	43.4	-43.4	-47.4	-60.1	-24.0	18.0
Growing	127	41.2	-24.3	-31.0	-81.0	44.7	62.8
Stable	6	40.9	-23.7	-26.0	-81.8	37.6	59.7
Unemployment Rate Growth from Start to Observation Year							
Declining	34	44.5	-28.9	-33.7	-88.3	41.3	64.8
Growing	67	40.2	-19.7	-26.5	-80.4	49.7	65.0
Stable	35	39.9	-30.0	-33.8	-73.0	40.3	56.6
HH Median Income Growth from Start to Observation Year							
Declining	43	35.6	-27.6	-28.1	-80.3	21.0	50.7
Growing	84	44.3	-22.8	-31.4	-82.8	44.8	63.8
Stable	9	39.6	-28.1	-34.9	-56.9	24.9	40.9
0 Vehicle HH Growth from Start to Observation Year							
Declining	28	48.3	-44.4	-48.7	-85.6	6.0	45.8
Growing	39	44.7	-16.3	-23.3	-79.3	50.9	65.1
Stable	69	36.3	-21.4	-26.5	-74.9	48.6	61.8
Gas Price Growth from Start to Observation Year							
Declining	51	42.7	-30.0	-31.8	-82.1	41.8	61.9
Growing	82	40.5	-22.2	-30.7	-78.6	45.2	61.9
Stable	3	34.7	-0.2	-9.7	-39.4	45.5	42.5

5.4.5 Ramp Up Period

Another effect on ridership forecast performance that have been evaluated in previous studies is the ridership ramp-up effect (Bent Flyvbjerg 2005; Chang et al. 2010; Shinn and Voulgaris 2019). Each year after project opening, ridership experiences a growth and therefore the deviation from forecasts get smaller. While we only considered a maximum ramp-up period of two years for our statistical analysis and model estimation, the database does allow us to explore the forecast accuracy for different ramp-up years. While the average deviation from forecast for these different ramp-up periods are different (Figure 24), the differences are not statistically significant as per the Tukey HSD Test. However, it should be noted that the spread of outcomes decreases with each additional year after project opening. It may have implications for the upper and lower quantiles of the uncertainty window.

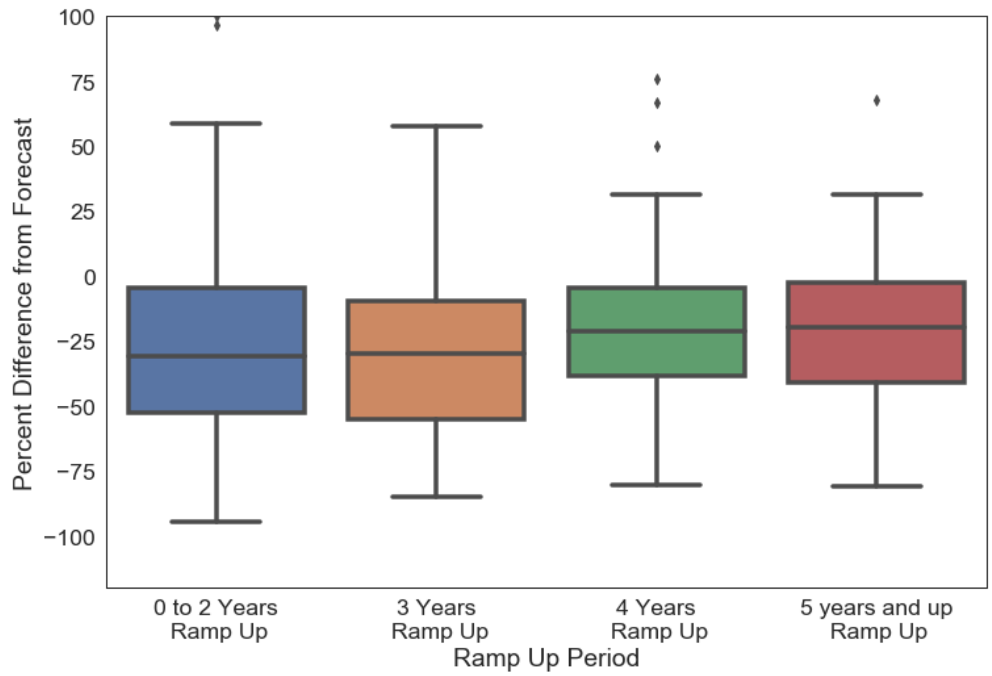

Figure 24: Transit Ridership Forecast Performance by Ramp Up Period

Factors that affect transit ridership, e.g., population and employment, zero-vehicle households and gas price, inevitably affects forecast performance. Metro areas experiencing stable growth (within $\pm 1\%$) in population, employment rate, household median income and gas prices had a smaller average deviation (MAPDF) from forecast than for ones experiencing greater change. Ridership increased with increasing zero-vehicle households and decreasing unemployment and household median income, resulting in smaller average deviation. The effect of changes in gas prices is not readily apparent. We can infer that a portion of accuracy in transit ridership forecasts can be attributed to the changes in such variables from project opening to observation.

Table 15: Effect of changes in socio-demographic variables during ramp-up on forecast performance

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
Metro Area Population Change During Ramp Up Period							
Declining	3	59.7	-25.4	-61.5	-65.7	40.2	53.0
Growing	90	43.3	-26.2	-31.9	-79.9	42.3	61.1
Stable	43	35.5	-21.3	-21.2	-83.3	43.1	63.2
Metro Area Household Median Income Change During Ramp Up Period							
Declining	39	38.1	-20.8	-25.6	-70.2	52.2	61.2
Growing	60	45.9	-26.3	-34.5	-81.1	44.2	62.7
Stable	37	36.9	-26.1	-28.1	-81.3	40.5	60.9
Metro Area Unemployment Change During Ramp Up Period							
Declining	82	43.7	-23.9	-31.1	-83.3	45.4	64.3
Growing	34	40.3	-29.7	-33.6	-73.9	39.1	56.5
Stable	20	32.4	-19.4	-24.8	-61.7	45.2	53.5
State Level Gas Price Change During Ramp Up Period							
Declining	62	41.0	-24.7	-23.9	-83.1	45.4	64.2
Growing	49	48.1	-28.2	-37.8	-83.6	39.4	61.5
Stable	25	28.2	-17.7	-20.0	-66.3	41.3	53.8
Metro Area 0 Vehicle Household Change During Ramp Up Period							
Declining	68	45.9	-24.5	-31.5	-77.3	43.3	60.3
Growing	34	36.5	-21.9	-25.6	-72.7	46.8	59.7
Stable	34	36.5	-27.7	-28.3	-85.5	33.2	59.3

5.4.6 Performance of the Capital Investment Grant Projects

The Capital Investment Grant (CIG) program by the FTA present a chance to compare methodological advances in ridership forecasting in the absence of more robust data on forecast methodology. The Before-After studies as part this program has led to several advances in the industry: improved methods for forecast, application of risk assessment methodology and maintaining proactive oversight of project operation (Federal Transit

Administration 2020). The difference in the mean PDF of CIG projects and non-CIG projects is statistically significant at 95% confidence level, with the non-CIG projects having a lower average deviation. About 85% of the CIG projects experienced ridership lower than the forecasts. More apt comparison of CIG and non-CIG project must take the year forecast was produced into account. Conforming to the Transportation Equity Act for the 21st Century (TEA-21) and FTA project assessment rule issued in 2000, the projects funded through this program required ex-post analysis of ridership and cost estimates (Transportation Research Board and National Academies of Sciences 2010). In 2001, FTA introduced new analytical tools which increased model scrutiny which may have resulted in better forecast performance. Ridership for projects that were produced after 2001 had mean PDF of -18.8% compared to -46% for the ones produced before this introduction. In

Figure 25 we present a more detailed breakdown of forecast performance by the year forecast was produced. The gradual improvement in performance is noticeable in projects funded through CIG programs and those that were not. The anomaly is the sole project forecasted in 2013 which had a much higher ridership than forecast. This project extends an existing line by 4 miles and had a forecast of only 2250.

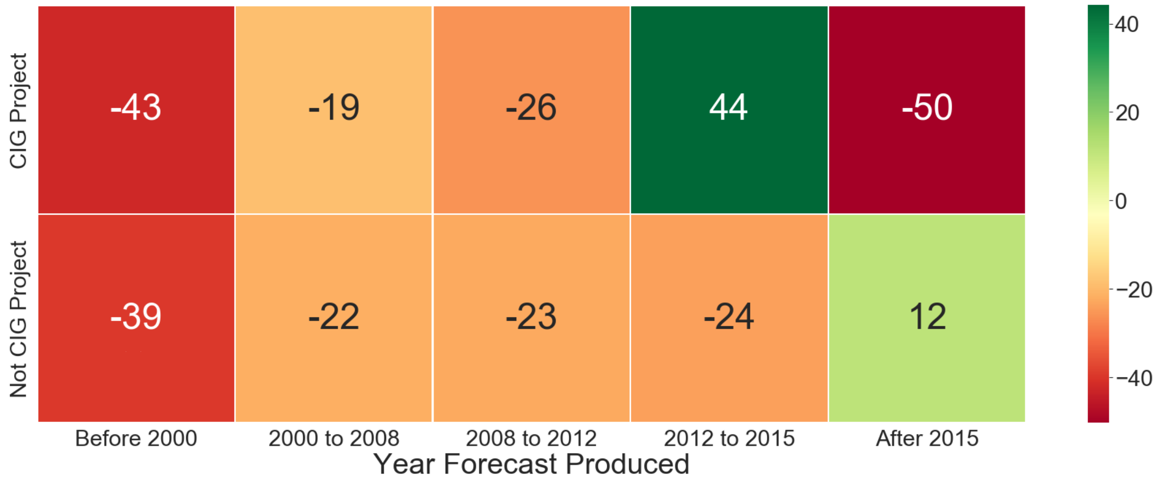

Figure 25: Forecast Performance for CIG and non-CIG Projects over the years (number of projects in parenthesis)

5.4.7 Forecast Horizon or Time Span

The greater the number of years between forecast production and measurement, the larger the opportunity for changes to have occurred in the economy, land use patterns, fuel prices, and other factors that influence travel. These are all variables that are difficult to predict, but their effects are evident. Our results show that as the time span increases, the forecasts get less accurate, with forecasts more than 5 years into the future having a statistically significant and larger average deviation. The absence of data on the year forecast was produced make other comparisons difficult.

5.4.8 Performance over the Years

In general, ridership forecasts for projects opening after 2000 show a noticeable improvement in performance (mean PDF of -22% over -52.9% before). Previous studies (D. Schmitt 2016; Voulgaris 2019a) have also noted this improved performance and attributes this to better scrutiny of the demand models in addition to improved technical methods. Figure 26 plots the 5-year rolling average of the mean PDF and the 95th and 5th percentile values from the sample. It is to be noted here that even though the average percent difference from forecast have been getting better, there remains significant spread of the outcomes.

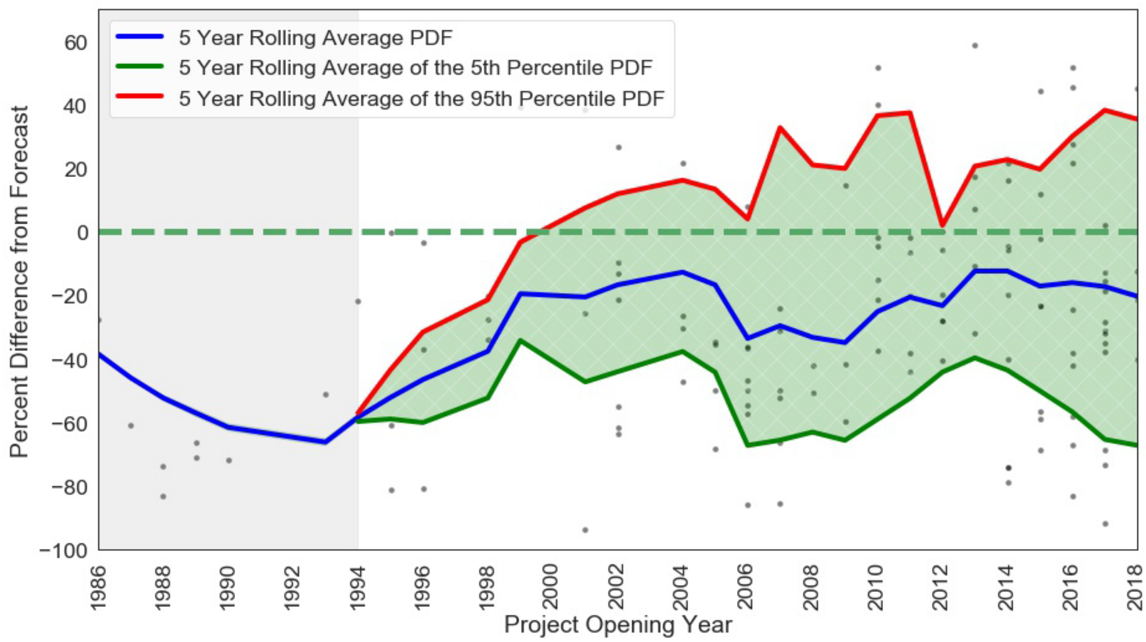

Figure 26: Transit Ridership Forecast Performance over the years

The plot of the rolling average PDF shows that the performance of ridership forecasts came to a halt in 2012 and started getting worse afterwards. This offset in performance can be attributed to the aggregate ridership trend as we explain in the following section.

5.5 Transit Ridership Trend and Forecast Accuracy

An important consideration while evaluating transit ridership forecast accuracy is the impact of overall transit trends— demand against the supply. We can quantify the demand by the unlinked passenger trips (UPT) and supply by vehicle revenue miles

(VRM). Figure 27 presents the transit supply (in vehicle revenue miles) and demand (unlinked passenger trips or ridership) in billions. We can see a positive correlation between UPT and VRM until the onset of the Great Recession. Even so, the demand and supply didn't increase at the same rate, leading to fewer ridership for every mile of service over the years. Projects that added service with the expectation of added ridership therefore didn't see it fulfilled. During the Great Recession years, supply dropped while the ridership kept increasing. The total ridership across all transit modes dropped after 2014, although the decline in ridership is apparent from 2012, especially for bus. Ridership changes relative to 2012 levels is presented in Figure 28.

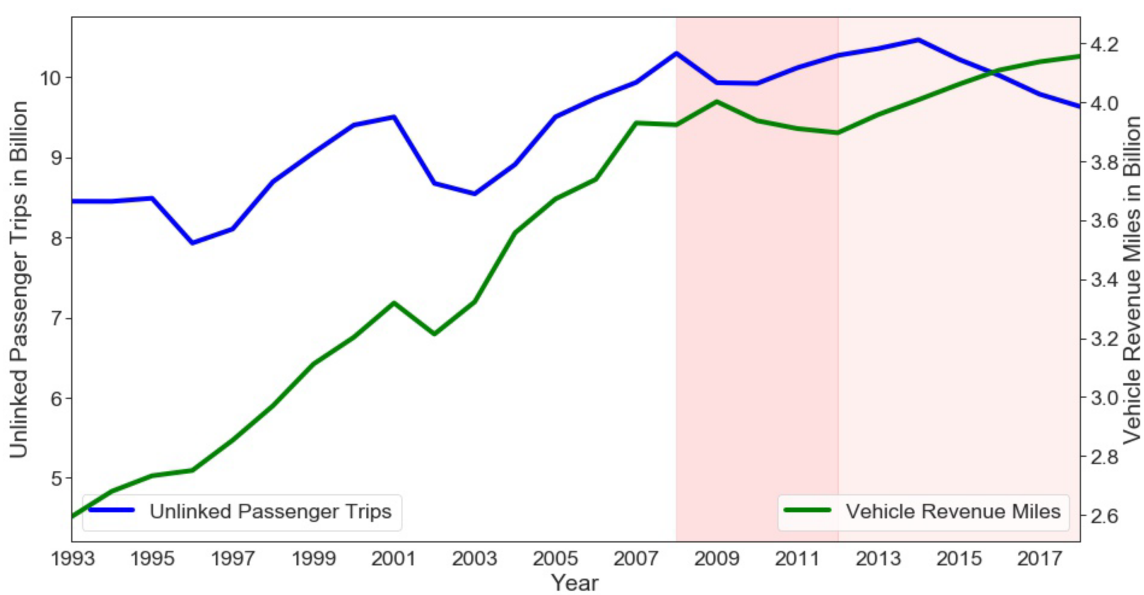

Figure 27: Transit Ridership Demand and Supply

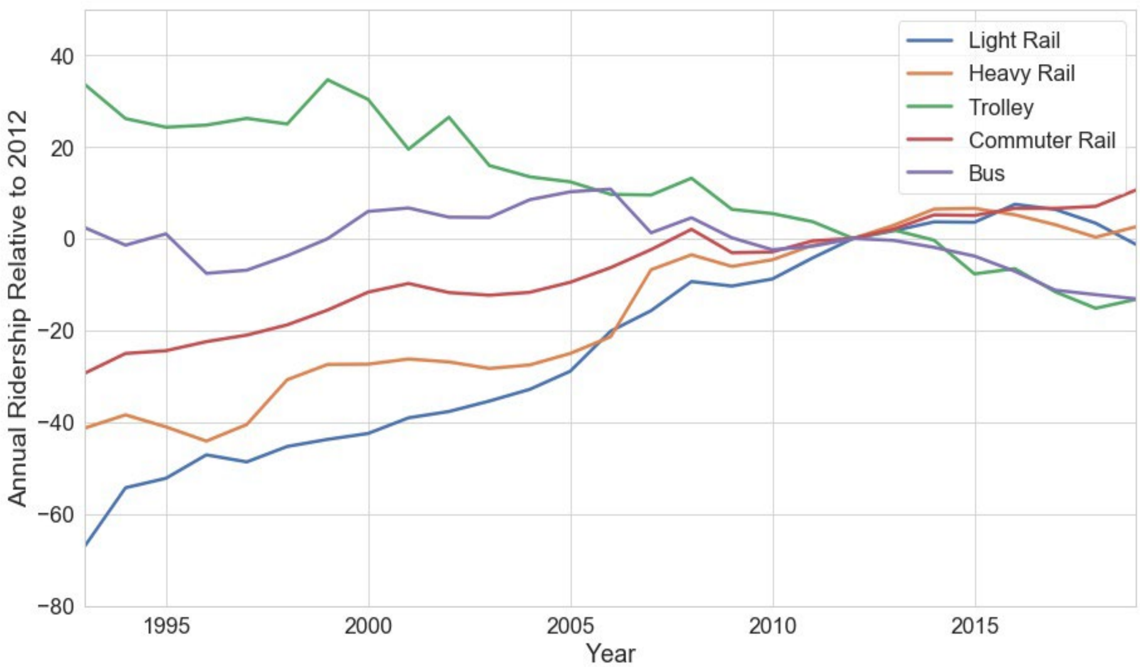

Figure 28 Annual Ridership Change relative to 2012 by Mode (Source APTA Ridership Report)

With changes in demand vs supply inevitably comes changes in ridership forecast performance, since forecasts are anticipated demand for changed supply. During the Great Recession when ridership was increasing despite decreasing supply, they got closer to the forecasts, resulting in smaller deviation from forecasts. After 2014, the opposite happened: the average deviation from forecasts got larger again. The trend is noticeable in Figure 29 which juxtaposes 5 year rolling average PDF and the demand vs supply curve. Forecasts produced during the Great Recession (2008 to 2012) have a higher mean PDF overall as well as higher than the mean PDF of each category of the ridership observation years. It is possible that forecasters may have overestimated ridership considering the high unemployment rate and car-ownership costs developed during the recession years. However, since 2012 to pre-COVID 2020, the US has enjoyed a steady and stable economy with low gas prices, resulting in fewer people than expected requiring transit. On the other hand, during the same period, there has been a growth in socio-economic demographics and land-use as well as improved transit services.

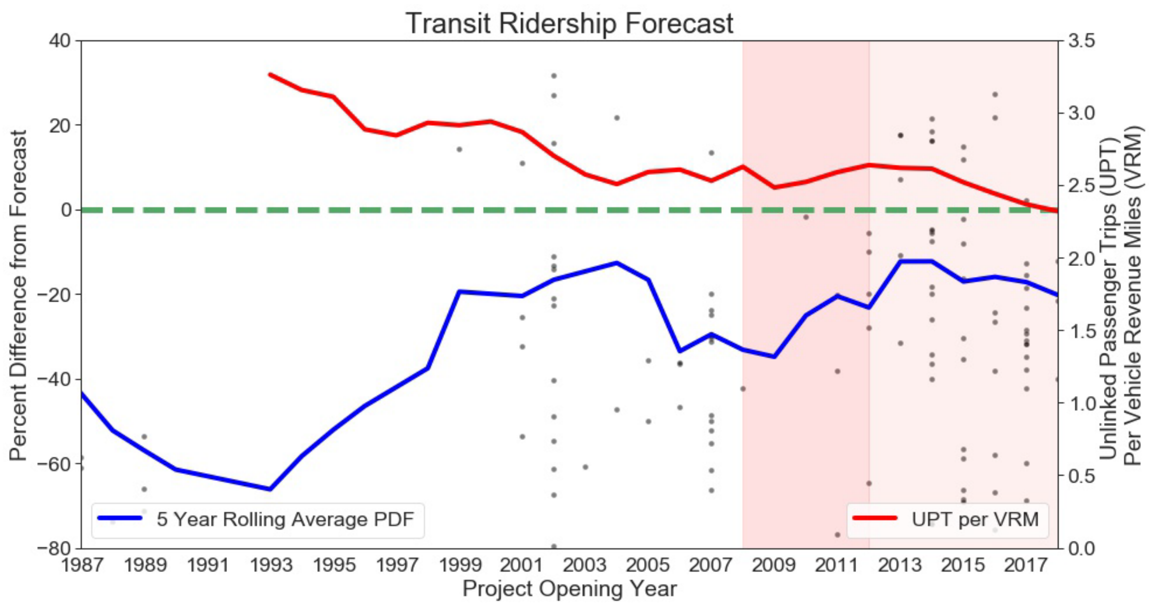

Figure 29: Transit Ridership Forecast Accuracy and Ridership Demand vs Supply

Performance of ridership forecasts over the years show that forecasts produced during 2000 to 2008 for projects opening during the Great Recession had a lower average deviation Figure 30. As ridership dropped unexpectedly from 2012, the PDF increased to -17%. Interestingly, the performance of forecasts produced during the Great Recession stayed relatively constant across the time.

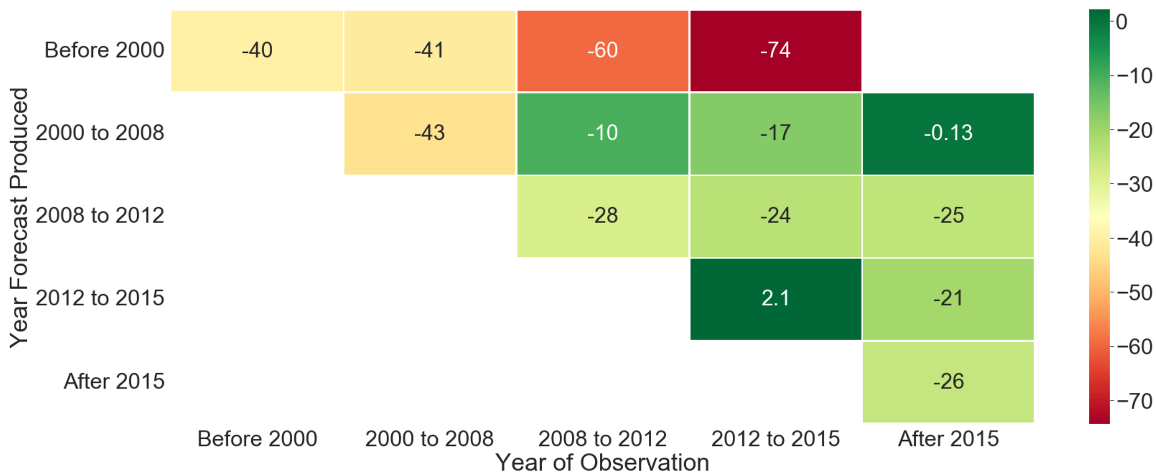

Figure 30 Mean PDF crosstabulation by Year of Observation against Year Forecast Produced (number of observations in parenthesis)

5.5.1 Adjusting for the Recent Decline in Transit Ridership

We identified a number of factors that affect transit ridership (G. D. Erhardt et al. Submitted), some of which result in increases and others in decreases to transit ridership. Together these factors result in a net bus ridership decline of 15% and a net rail ridership decline of 3% between 2012 and 2018. While several factors contribute to lower transit ridership, including lower gas prices, higher fares, and changes to income, teleworking rates, and car ownership, we show that ride-hailing is the most important. By 2018, ride-hailing reduced bus ridership by 10% and reduced rail ridership in mid-sized metropolitan areas by a similar amount. It had a positive, but insignificant effect on rail ridership in the largest metropolitan areas.

- Transit connects people to activities and jobs, so the number and location of both affect transit ridership. Each 1% increase in population plus employment is associated with 0.22% more transit ridership. Similarly, higher density leads to more transit ridership. We considered the percent of the population and employment in a region that is within a transit supportive density, defined as more than 10 people or employees per acre. For each percentage point increase (such as from 10% to 11%) in population plus employment living in these denser areas, transit ridership is 0.4% higher.

- Higher gas prices make driving more expensive and incentivize travelers to use transit. Each percent increase in gas price accounts for a 0.14% increase in transit ridership.
- Several factors related to the characteristics of households, their income, and their work norms may affect transit ridership. We find three to be important: income level, 0-vehicle households, and telecommuting. With higher per capita income, people are less likely to ride transit. Each 1% increase in the median per capita income results in a 0.07% decrease in transit ridership.
- Higher shares of 0-vehicle households in an MSA have a small positive effect on transit ridership. We know that people from households without a car constitute an important market of transit riders. However, our estimated coefficient is small, so the results show that the change in vehicle ownership explains little about the change in transit ridership over this period. Our results show that a decrease from 10% of households owning 0 vehicles to 9% of households owning zero vehicles would result in 0.2% less transit ridership, but this effect is not statistically significant.

As unemployment, percent of zero-vehicle households and average gas prices increase in a metro area, people tend to use transit more. For optimistically biased forecasts, this means that actual ridership would be closer to the forecast which is what we see in our analysis. Ridership on projects that opened after 2012 was about 18.7% lower than forecast on average. After adjusting the ridership for metro area population, employment, household income, zero vehicle households, presence of TNCs and changes in gas prices from project opening to observation, as well as the type of project (maintenance and/or network restructure), this PDF comes down to -9.4%. The absolute deviation decreases as well.

Table 16: Transit Ridership Forecast Accuracy if actual ridership is adjusted for the decline

	Obs.	MAPDF	Mean	Median	5th Percentile	95th Percentile	(95th - 5th Percentile) / 2
Overall Distribution							
Total Sample	136	41.2%	-24.7%	-30.7%	-81.6%	45%	63.3%
Adjusted Sample	123	40.5	-21.3	-27.8	-80.7	51.2	65.9
Subset of Projects Opening On or After 2012							
Original Sample	65	37.4	-18.9	-23.3	-78.3	50.4	64.4
Adjusted Sample	52	34.9	-9.4	-15	-75.4	68.7	72.1

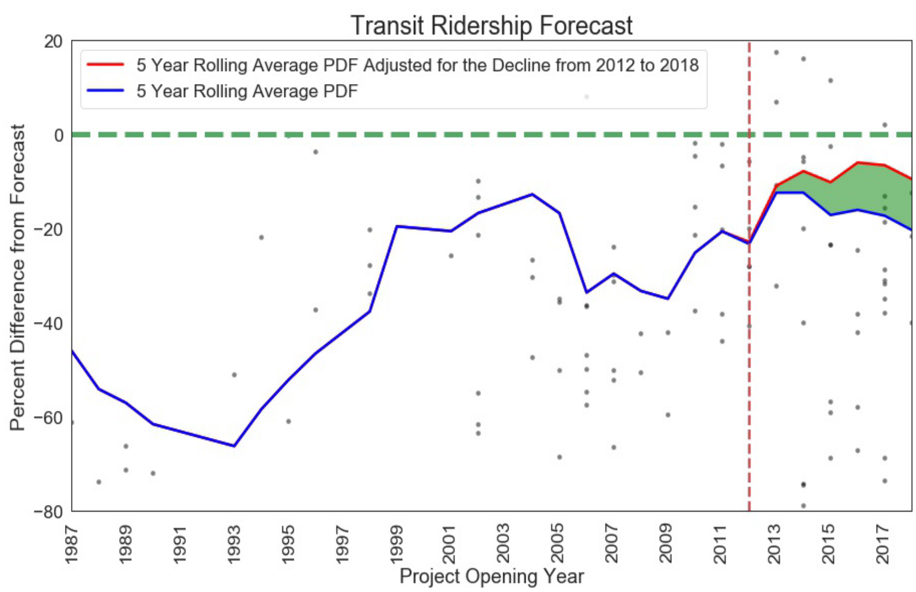

Figure 31: Transit Ridership Forecast Performance after adjusting for the decline since 2012

5.6 Modelling forecast accuracy

We tested a series of linear regression models to test the sensitivity of ridership forecast accuracy against the explanatory variables presented in Section 5.4.While traditional methodology dictates the use of Ordinary Least Squares regression, we instead chose to model accuracy using Quantile Regression formulation. Several key differences exist between the two methods; the most important one is the assumptions of regression. OLS estimates the differences in the outcome variables at the mean after adjusting for other explanatory variables. This assumes that the regression coefficients are constant across the population. In contrast, the QR method weighs the distances between the values predicted by the regression line and the observed values, and minimizes the weighted distances (Lê Cook and Manning 2013). For this reason, QR method relaxes the assumption of normality of the error term. Moreover, this technique is robust to the presence of outliers (Barnes and Hughes 2002).

To model forecast accuracy, we estimated quantile regression models for the 50th or median quantile using the following framework:

$$y_{P,i} = \alpha_P + \beta_P \hat{y}_i + \delta_P X_i \hat{y}_i + \varepsilon_{P,i} \qquad \textit{Equation } 4$$

Where y_i is the counted traffic on project i, $\hat{y}_i$ is the forecast traffic on project i and ε_i is a random error term. α and β are estimated regression coefficients, while X_i is a vector of descriptive variables associated with project i, and δ is a vector of estimated model coefficients associated with those descriptive variables. The P index indicates that the term applies to the 50th percentile. The coefficients of the models are estimated by minimizing the weighted sum of absolute error. The weights in the minimization function are themselves dependent on the quantiles of interest. With these models we can detect the effect of regressors on the median expected value, so that 50% of the observation fall on either side of the regression line. For example, consider a model where α is 0, β is 1 and there is a single descriptive variable, $X_{1,i}$, which is a binary flag which is 1 if the forecast serves the CBD of the metro area, and 0 otherwise. If γ_1 has a value of -0.1 then it means that the median actual value would be 10% lower than the forecast. If γ_1 has a value of +0.1 then it means that the median actual value would be 10% higher than the forecast.

5.6.1 Model Estimation Results

In this section, we present the estimation results for two models (Table 17): the first one is a simple median quantile model of observed ridership against forecast to establish the overall trend of accuracy. The second one includes the explanatory variables by project and area characteristics. It is to be noted that the changes in external variables from the project start year is often unavailable due to the lack of data on project starting. Moreover, the variables themselves interact among themselves in such a way that they sometimes change the direction of the coefficients. The full model presented below contains the variables that make sense in addition to producing better goodness-of-fit measure and consistent coefficients across the descriptive variables. The greyed cells represent the variables that are not significant at a 90% confidence level but kept in the final model because they produce a better fit and have a logical interpretation. The interaction of these variables has been tested, but the interpretation is not clear.

Table 17: Transit Ridership Forecast Accuracy Model Estimation Results

Pseudo R-Squared	**Simple Model**		**Full Model**	
	0.61		**0.82**	
Variable	**Coeff.**	**(t value)**	**Coeff.**	**(t value)**
Overall Distribution				
Intercept (α)	1709.25	2.02	1153.99	1.49
Forecast Ridership (β)*	0.5	4.54	0.52	3.00
Mode (Urban Heavy and Light Rail as reference, total 58 observations)				
Busway (2 observations)			-0.23	-1.84
Bus Rapid Transit (32 observations)			-0.32	-2.37
Commuter Rail (22 observations)			-0.20	-2.01
Streetcar (18 observations)			-0.86	-2.86
Project Service Area (outside the CBD as reference)				
Serving the Central Business District of the Metro Area			0.23	2.54
Project Opening Year (Before 2005 as reference)				
Number of years after 2005			0.04	1.59
Number of years after 2012			-0.02	-0.30
Ramp Up Period				
Number of years after project opening			0.04	0.64
Time Span				
Number of years between start and opening			-0.03	-2.03

We discuss the variables below:

Overall Distribution

The overall distribution includes the intercept (α) and the forecast volume (β). We can think of these values as a reference line, with the remaining terms (γ) in the model changing the slope of that reference line. The median transit ridership is about 50% of the forecast on both simple and full model, confirming earlier observation (in this research and the previous studies) of the presence of optimism bias.

Project Mode

Forecasts on the urban light and heavy rails have performed better than the rest. Compared to them, bus, BRT, commuter rail and streetcars have had lower average weekday riders. Since light and heavy rails are taken as the reference, the coefficients on the other modes represent the additional deviation from forecast compared to the reference.

Project Service Area

Projects that serve the central business district have higher weekday boardings, and therefore the deviation from forecast is less than the ones that serve outside the CBD. It is generally assumed that work travel patterns are easier to model than non-work travel because of the publicly available home-to-work records in the American Community Survey and Longitudinal Employee-Household Dynamic (LEHD) data. Since the travel pattern in the CBDs are relatively easier to predict, it is associated with greater accuracy.

Project Opening Year

As we have surmised from our categorical exploration, forecasts have been getting better over the years. We tested it using several breakpoints and project characteristics: project forecasted before and after 2001 when new evaluation tools have been introduced by the FTA, start year between 2008-2012 (the Great Recession years), and project opening after 2005. Since we have a lot of missing data on the year forecast was produced, in the final model we used project opening year as a substitute. Projects open several years after they have been forecasted for; most of the projects that opened on or after 2005 were forecasted after 2000, the year associated with updated methodology. Each year after 2005, ridership got 4% higher which led to lower percent deviation from forecasts.

However, this positive effect is partially offset by the recent transit ridership trend. Our analysis of transit ridership decline during this period reveal that this decline is caused by the presence of TNCs, declining gas prices, decreasing zero vehicle households, and increasing household income. Effects of these variables are difficult to estimate for ridership forecasts, because of the cross-sectional nature of the data. We can infer from the substitute of years after 2012 that each year saw lower average ridership and higher deviation from forecasts. Ridership for a transit project opening in 2018 will therefore increase by 52% (0.04X13) for 13 years after 2005, decrease by 12% (0.02X6) for 6 years after 2012, resulting in a 40% increase in ridership. All else remaining equal, ridership would be about 92% of the forecast, if all else remain equal. These effects are not statistically significant, but they point to forecasts getting better over the years and external factors influencing forecast performance.

Ramp Up Period

Each year after project opening, ridership experiences a growth and therefore the deviation from forecasts get smaller. While we only considered a maximum ramp-up period of two years for our statistical analysis and model estimation, the database does allow us to explore the forecast accuracy for different ramp-up years. The effect of the ramp-up period is positive on observed ridership, thereby improving the forecast accuracy.

Time Span

The time span measures the years between when a forecast is made and the opening-year. Time span coefficients are positive and significant for all percentiles, meaning that the post-opening traffic volume is more likely to exceed a long-term forecast than a short-term forecast. This result may reflect overall traffic growth rates exceeding expectations and slowly building over time. However, longer-term forecasts are likely to be associated with larger projects, so distinguishing the potential effects of project size and time span remains difficult.

5.7 Summary of Findings

In this paper, we present the evidence of persisting optimism bias in transit ridership forecasts. Transit ridership in the projects in our database are about 24.6% lower than the forecasts on average, with an average deviation of 40.5%. About 90% of the projects in the database had observed ridership within -81.6% to 45.2% of the forecast ridership. Below we present some of the key findings of this analysis:

- **Transit ridership forecasts are optimistically biased**: We found that about 70% of the projects in our database had ridership lower than the forecast. An implication of these results is that decision makers and the public are likely provided with inaccurate and optimistically biased information and, assuming this information is used to support a decision favorable to constructing the transit project, advancing projects that might would otherwise be not funded or revised to be made more efficient.
- **Forecast accuracy is getting better, but uncertainty remains integral**: Transit projects that have opened after 2000 has seen ridership much closer to the forecast than before. It can be attributed to better models, better data and even organized ex-post evaluation program that incentivizes good forecasts. However, significant spread remains in the outcomes against

forecast. Similar to traffic forecast, this spread points to the need of incorporating uncertainty in ridership forecast for effectively policy planning.

- **Scope of the service provided by the modes may contribute to the inaccuracy**: Forecasts for bus rapid transit, urban light rail and streetcar or trolley perform better on average than heavy and commuter rails. These two modes typically serve longer routes with heavier traffic than light rail, streetcars, and people movers. The travel models used to forecast the ridership may not adequately account for the large network with high variability in demand in the analysis. Projects with a smaller length mean fewer stations and fewer ridership, in addition to less sensitivity to land-use and economic changes. Because of their length, streetcars and trolleys have a smaller scope and more accurate forecasts. Commuter rails, on the other hand, typically serve longer distances and therefore have a much larger scope contributing to more degrees of freedom.
- **Institutionalized review programs help getting better forecasts**: The Capital Investment Grant (CIG) program by the FTA conducts before-after studies as an integral element to project funding. This has increased scrutiny of the forecasts and contributed to several advances in the forecasting methodology. In 2001, FTA introduced new analytical tools which increased model scrutiny which may have resulted in better forecast performance. Ridership for CIG projects that were produced after 2001 had mean PDF of -17.4% compared to -43.4% for the ones produced before this

introduction. Even the non-CIG funded projects produced after 2001 performed better than the ones produced before (-19.7% against -39.3%), although the sample size and data availability is a barrier to statistical significance test.

- **Transit ridership forecasts vary by their location and area transit characteristics**. We found ridership for projects that are in the Central Business District (CBD) of the metro area has greater over-prediction than those outside. One possible explanation can be the dependence of ridership forecasts on employment, which is typically reflected in the CBD. Another important factor of transit ridership is the area's familiarity with transit systems. Project sponsors serving larger populations may have greater resources to devote to preparing rigorous forecasts. They may also answer to a wider variety of stakeholders, which could influence the incentives for promoting a particular project through optimistic forecasts.
- **Auto ownership and gas prices affect forecast performance.** Transit ridership is dependent on the unemployment rate, percent of zero-vehicle households and average gas price of the area as well. Prior research works have established that growth of such economic factors is typically associated with higher transit ridership. In our analysis we found that the metropolitan statistical areas experiencing growing unemployment rates, zero vehicle households and increasing gas prices from the start year to the observation saw lower average deviation from forecasts. As transit ridership grew in these metro areas, the deviation from forecasts got low as well.

Ridership on projects that opened after 2012 was about 18.7% lower than forecast on average. After adjusting the ridership for metro area population, employment, household income, zero vehicle households and presence of TNCs, this PDF comes down to -12.2%. The absolute deviation decreases as well.

While transit ridership forecast accuracy has improved, there remains substantial deviation in the observed ridership from forecast ridership. It is prudent that this uncertainty is acknowledged in the forecasts themselves by presenting a range of values rather than a point forecast using scenario analysis or sensitivity tests. The quantile regression method described in the next chapter provides an alternate method to constructing an uncertainty window around forecasts, aside from establishing the effect of different factors on the accuracy.

Chapter 6 ESTIMATING THE UNCERTAINTY IN TRAFFIC AND TRANSIT RIDERSHIP FORECASTS

The evidence of uncertainty around travel demand forecasts as presented in the previous chapters point to the necessity of expressing the forecasts as a range of expected outcomes. Traditional methods for estimating such uncertainty windows rely on assumptions about reasonable ranges of travel demand forecasting model inputs and parameters. Rather than relying on assumptions, we demonstrate how to use empirical measures of past forecast accuracy to estimate the uncertainty in future forecasts. We develop an econometric framework based on quantile regression to estimate an expected (median) volume as a function of the forecast, and a range within which we expect 90% of volumes to fall. Using data on observed versus forecast traffic for 1,291 road projects, we apply this framework to estimate a model of overall uncertainty and a full model that considers the effect of project attributes. Our results show that the median post-opening traffic is 6% lower than forecast. The expected range of outcomes varies significantly with the forecast volume, the forecast method, the project type, the functional class, the time span, and the unemployment rate at the time forecast is made. Similarly, the uncertainty around transit ridership forecasts is determined by their mode and coverage, methodological advances after 2005, ridership trend since 2012, mode, years after opening and time span. A forecaster can apply the resulting equations to calculate an uncertainty window for their project, or they can estimate new quantile regression equations from locally collected forecast accuracy data. Aided by decision intervals, such uncertainty windows can help planners determine whether a forecast deviation would change a project decision.

This chapter has been adapted from, and extends to include transit ridership forecast uncertainty to the following published paper:

Hoque, J.M., Erhardt, G.D., Schmitt, D., Chen, M. and Wachs, M., 2021. Estimating the uncertainty of traffic forecasts from their historical accuracy. *Transportation research part A: policy and practice*, *147*, pp.339-349.

6.1 Introduction

Travel demand forecasting informs decisions about transportation projects. Good forecasts should provide valuable information to aid decision making (Murphy 1993) and they should be accurate to the point that different forecasts would not change the decision (Voulgaris 2019b). Several authors have advocated for using uncertainty windows that provide a range of forecasts (Bain 2011; Hartgen 2013), while Anam, Miller, and Amanin (2020) demonstrate the use of decision intervals to determine whether a forecast error would change a project decision. If it would not, then the sponsor can safely proceed with the project with respect to forecasting risk. If a value within the range could lead to a different decision, the sponsor may consider further study to better understand the risks involved.

The quantifiable uncertainty in travel demand forecasts primarily results from two sources: model inputs and the models themselves in their specification and parameters (Hugosson 2005). Model inputs include exogenous variables like assumptions about the completion of other projects in the transportation network, fuel prices, and socio-demographic and economic projections. These inputs are uncertain quantities themselves and susceptible to disruptions such as the Great Recession and the COVID-19 pandemic. The two sources are not mutually exclusive as well, because input errors propagate through travel demand models resulting in forecast error (Zhao and Kockelman 2002).

We completed this work as part of a National Cooperative Highway Research Program (NCHRP) project on traffic forecast accuracy. NCHRP Report 934 (G. Erhardt et al. 2020) provides additional details, including forecast accuracy metrics, case studies, and recommendations for improving traffic forecasting methods by systematically evaluating their accuracy.

6.2 Method

We begin by defining the conceptual relationship between accuracy and uncertainty. Then we present an econometric framework to measure accuracy and estimate uncertainty windows.

6.2.1 The Conceptual Relationship between Traffic Forecast Accuracy and Uncertainty

As Figure 32 illustrates, accuracy and uncertainty are deeply intertwined concepts, especially in the context of forecasting in planning. Accuracy is the closeness of a measurement or estimate to its true value (ISO 5725-1 1994). Uncertainty is the range in which a true value lies with some level of confidence (ISO/IEC Guide 98-3 2008). In forecasting, we treat post-opening traffic counts as an observation of the true value with the caveat that the counts themselves are subject to measurement error. Evaluating accuracy is a retrospective activity that accounts for past forecast errors, while expressing uncertainty is a prospective activity that considers possible errors. Because an uncertainty estimate is a "*means of expressing the accuracy of results*" (Collaboration for Nondestructive Testing n.d.), we should consider observations of historical accuracy when estimating uncertainty windows. We propose that the comparison of observed versus forecast traffic for past projects should be used to estimate the range of possible traffic volumes in future forecasts.

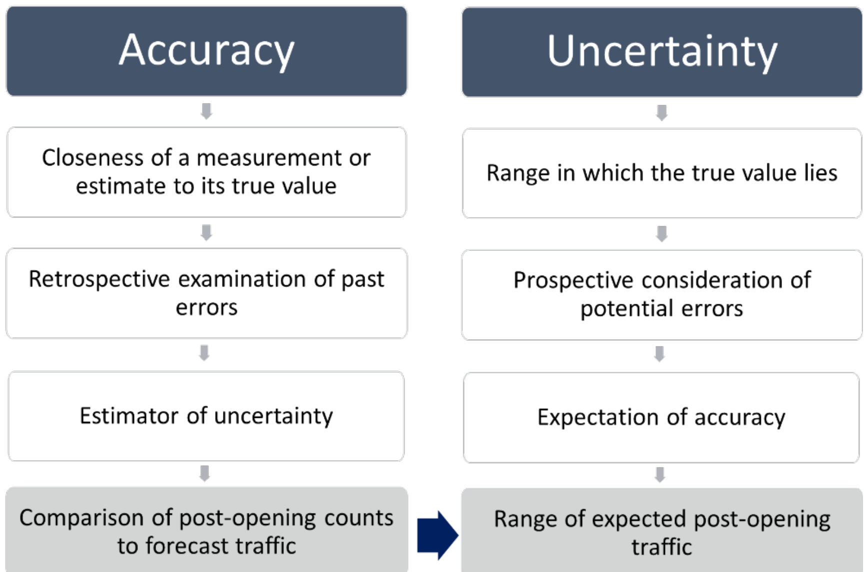

Figure 32: Relationship between forecast accuracy and uncertainty

6.2.2 Econometric Framework

We started from an econometric framework proposed by Odeck and Welde (2017). They regress the counted volume as a function of the forecast value using the equation:

$$y_i = \alpha + \beta \hat{y}_i + \varepsilon_i \qquad \textit{Equation } 1$$

where y_i is the counted traffic on project i, $\hat{y}_i$ is the forecast traffic on project i and ε_i is a random error term. α and β are estimated regression coefficients. Odeck and Welde suggest rejecting the null hypothesis that the forecasts are unbiased if α is significantly different from 0 or β is significantly different from 1. Starting from this structure, we introduced additional terms as descriptive variables:

$$y_i = \alpha + \beta \hat{y}_i + \delta X_i + \varepsilon_i \qquad \textit{Equation } 2$$

where X_i is a vector of descriptive variables associated with project i, and δ is a vector of estimated model coefficients associated with those descriptive variables. To consider multiplicative effects rather than additive effects, we multiplied the regressors by the forecast volume:

$$y_i = \alpha + \beta \hat{y}_i + \delta_i X_i \hat{y}_i + \varepsilon_i \qquad \textit{Equation } 3$$

In this formulation, δ =0 indicates no effect of that term, while positive values would scale up the forecast and negative values would scale down the forecast. The coefficients of categorical variables signify their effect compared to an omitted reference level. For example, consider a model in which α is 0, β is 1 and there is a single descriptive variable, $X_{1,i}$, a binary flag which is 1 if the forecast is for a new road, and 0 for a project on an existing roadway. If δ_1 has a value of -0.1 the expected value would be 10% lower than the forecast. If δ_1 has a value of +0.1 the traffic count would be 10% higher than the forecast.

With the above formulation we can explore the variables associated with higher or lower traffic relative to forecast but can say nothing of the distribution beyond the mean. For example, forecasts with longer time horizons may be no higher or lower on average but may have a wider range of outcomes. Therefore, we extend the above framework to use quantile regression instead of ordinary least square (OLS) regression. Whereas OLS predicts the mean value, quantile regression predicts the values for specific percentiles in the distribution (Cade and Noon 2003). In addition, Quantile Regression Methodology does not assume any parametric distribution (e.g. normal, Poisson etc.) of the random error term in the model, unlike OLS. Zhang and Chen (2019) used quantile regression to quantify the effect of weather on travel time reliability, where an event may have a small effect on the mean value but increase the likelihood of a long delay. In an application analogous to this project, Pereira et al. (2014) used quantile regression to estimate error bounds for real time traffic predictions.

We estimated quantile regression models of counted traffic as a function of the forecast and other descriptive variables. We did so for the 5th percentile, the median, and the 95th percentile, with separate regression equations for each:

$$y_{P,i} = \alpha_P + \beta_P \hat{y}_i + \delta_P X_i \hat{y}_i + \varepsilon_{P,i} \qquad \text{Equation 4}$$

where the P index indicates that the term applies to the 5th, 50th or 95th percentile. The coefficients of the models are estimated by minimizing the weighted sum of absolute error. The weights in the minimization function are themselves dependent on the quantiles of interest. With these models we can detect the effect of regressors on the median expected value and on the range of outcomes. Variables with positive coefficients in the 5th percentile model and negative coefficients in the 95th percentile model indicate a narrower uncertainty window.

6.3 Data

Estimating the proposed models requires a sufficiently large sample of forecasts and post-opening observations. For this study, we used a data set of forecasts and counted Average Daily Traffic (ADT) on 1,291 road projects in the United States and Europe, including new roads, capacity expansion projects, operational improvements and resurfacing projects. We compiled these data for NCHRP 934 (G. Erhardt et al. 2020) and processed them as described in (Hoque, Erhardt, Schmitt, Chen, Chaudhary, et al. 2021b). The data also include project attributes, as Table 18 shows.

Consistent with Flyvbjerg's (2005) recommendations, we evaluated opening-year conditions, which we defined as the first post-opening year for which we have counts. Sometimes a project is delayed, and the opening-year differs from what was previously expected. In those cases, we scaled the forecast traffic to the opening-year using the growth rate implied by the opening and design year forecasts (usually 20 years after projected opening) where it is available. A standard traffic growth rate of 1.5% was assumed by (Nicolaisen 2012) in his analysis of forecast accuracy of European projects, also a part of our analysis, and we carried the convention forward in our dataset.

Table 18: Data Fields

Variable Name	**Description**	**Data Availability**
Forecast	Forecast daily traffic.	100%
Count	Counted daily traffic.	100%
Agency Type	Variable describing the type of the agency producing the forecast—State DOT, MPO or consultant	56%
Agency	Geographic location of project by State/Country. Corresponds to the origin of the datasets included in the database.	100%
Functional Class	US Federal Highway Administration specified functional classification of the roadway.	72%
Area Type	The area type where the facility lies: Rural, Mostly Rural, Urban and Unknown area types according to US Census Bureau's definition of Urban and Rural areas. The Bureau defines urban areas as a territory that has at least 2,500 people. The percentage of people living in rural areas in a county determines whether the county is rural (100%), mostly rural (50-99%) or urban (<50%).	91%

Improvement Type	Type of project: improvement on an existing roadway, new construction project.	72%
Forecast Method	Methodology for forecasting: using travel demand model, population growth rate, traffic count trend, professional judgement.	48%
Start Year	The year when forecast was produced.	100%
Forecast Year	The year forecast was produced for, usually opening year, interim year (usually 10 years after opening and design year (20 years after opening).	100%
Opening Year	The earliest year after project opening that traffic count data are available.	100%
Time Span	Number of years between year forecast was produced and forecast year.	100%
Unemployment Rate	State level unemployment rate in the start year, forecast year and opening year, obtained from the Bureau of Labor Statistics. For European projects, the national unemployment rate was obtained from the World Bank historical unemployment rate data	100%

Our database compiles projects opening from 1970 to 2017.Most agencies that systematically track traffic forecasts have only begun the practice within the past 10 to 15 years. As a result, 90% of projects in our data opened in 2003 or later. Routine projects such as repaving and minor improvements are more common in more recent years, as agencies are less likely to maintain records of those projects over a span of decades. Practices vary for which project attributes to record with a forecast, so attributes such as the type of project and the forecasting method used are often missing.

The 1,291 projects are comprised of 3,912 individual road segments with both forecasts and post-opening counts. The estimation can be done in two ways— considering each segment as a separate observation or aggregating across projects to get the average forecast and average count. Considering each segment as a separate observation may introduce a bias since the segments in a project are correlated among themselves. However, aggregating across projects would remove the diversity in roadway functional classes and area types (for projects spanning across several types of roadways) and volume (major and minor approaches to intersections). In an analysis of stock market returns, (Barnes and Hughes 2002) argue that aggregation results in imprecise estimates particularly at the extreme quantiles by "diversifying away the effect of individual observations or the impact of the omitted variables in the model". We tested a set of models that aggregated the results to the project level, then estimated quantile regression models and found that the project-level models produced narrower uncertainty windows. We therefore consider each segment as unique observation in our analysis.

6.4 Traffic Forecast Uncertainty Model Estimation

In this section, we present the results of two quantile regression models following the framework above. We estimated both using the `quantreg` package in `R` (Koenker et al. 2018). The first model, which we refer to as the Base model, indicates the overall uncertainty window. The second model, referred to as the full model, includes additional exogenous regressors. Later in this paper, we demonstrate how to apply the full model to calculate uncertainty windows. In discussing these results, we refer to the prediction of the quantile regression models as the expected traffic volume, which is distinct from the forecast traffic volume that is treated as an input to the quantile regressions.

6.4.1 Base Model

Table 19 presents the regression statistics for the base model. The gray cells indicate variables that are not statistically significant at the 95% confidence interval. While α_{50th} is not significantly different from zero, β_{50th} is significantly different from one, indicating a detectable shift. In this case, the median expected traffic volume is about 6%

lower than forecast. Figure 33 plots the counted versus forecast ADT and all three quantile regression lines. We consider the area between the 5th and 95th percentile regression lines to be the uncertainty window—the range within which we expect 90% of counted traffic volumes to fall. When α_{5th} and α_{95th} are close to 0, and β_{5th} and β_{95th} are close to 1, it indicates more accurate past forecasts and a narrower uncertainty window.

Table 19: Quantile regression results for base model of daily traffic volume

	5th Percentile		**50th Percentile**		**95th Percentile**	
Pseudo R-Squared	0.433		0.723		0.748	
	Coef.	(t value)	Coef.	(t value)	Coef.	(t value)
Overall Distribution						
Intercept (α)	-826.73	(-10.55)	37.15	(0.54)	2940.45	(6.50)
Forecast Volume (β)*	0.624	(-18.43)	0.941	(-9.28)	1.421	(12.52)

* t values for Forecast Volume are relative to 1 not 0.

The results in equation format following Equation 4:

$$y_{5,i} = -826.73 + 0.624 * \hat{y}_i + \varepsilon_{P,i}$$

$$y_{50,i} = 37.15 + 0.941 * \hat{y}_i + \varepsilon_{P,i}$$

$$y_{95,i} = 2940.45 + 1.421 * \hat{y}_i + \varepsilon_{P,i}$$

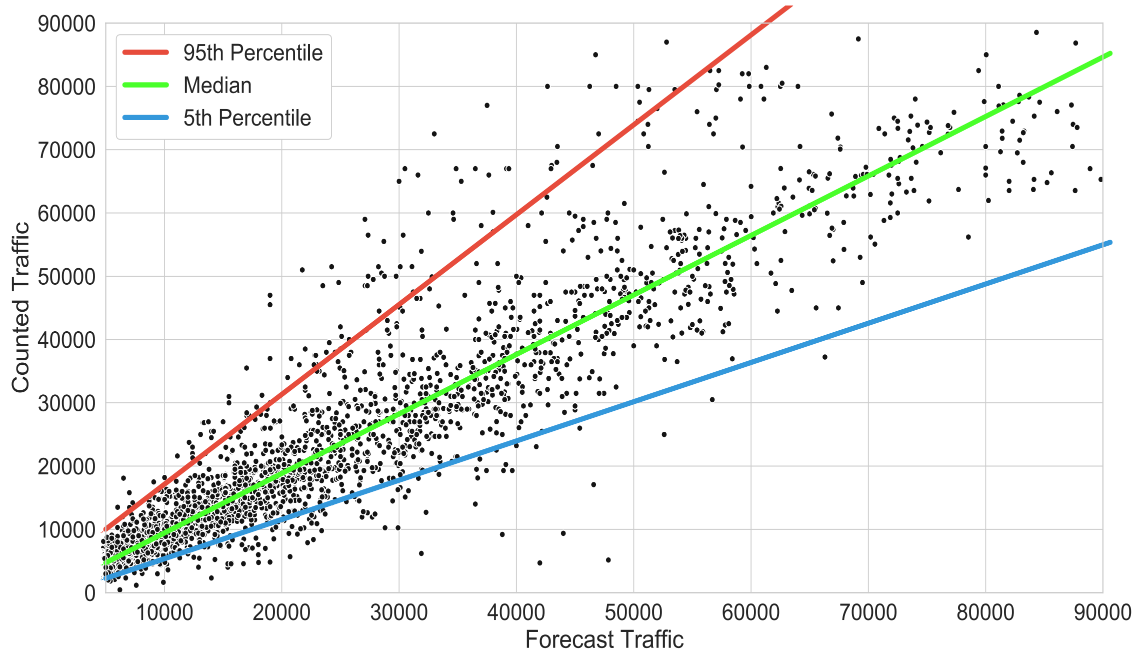

Figure 33: Counted versus forecast traffic and base model quantile regression lines

6.4.2 Full Model

Table 20 shows the model estimation results for the full model. Except for those in the overall distribution group, we interacted all terms with the forecast volume such that they served as scaling factors. We indicate the reference group for categorical variables above the relevant rows. We describe the interpretation of the coefficients below.

Table 20: Quantile regression results for full model of daily traffic volume

	5th Percentile		50th Percentile		95th Percentile	
Pseudo R-Squared	0.475		0.739		0.830	
	Coef.	t value	Coef.	t value	Coef.	t value
Overall Distribution						
Intercept (α)	-182.26	(-1.77)	255.55	(4.67)	976.78	(4.79)
Forecast Volume (β)*	0.705	(-6.69)	0.891	(-5.53)	1.254	(4.83)
Forecast Volume in excess of 30,000 ADT	0.024	(0.57)	-0.004	(-0.22)	-0.413	(-9.89)
Time Span						

Time span (years)	0.006	(2.81)	0.008	(5.62)	0.020	(10.50)
Unemployment Rate						
Unemployment rate in the year forecast was produced (%)	-0.006	(-1.41)	0.002	(0.87)	0.010	(1.87)
Binary Variables						
Functional Class (Reference class = Freeways)						
Major or minor arterials	-0.150	(-5.24)	-0.062	(-5.17)	-0.116	(-5.88)
Collectors and local roads	-0.212	(-4.03)	-0.126	(-5.21)	-0.321	(-2.36)
Project Type (Reference class = Existing Road)						
New road	0.093	(4.34)	-0.008	(-0.90)	-0.090	(-4.29)
Forecast Method (Reference class = traffic count trend, population growth rate, or professional judgment)						
Travel demand model	0.068	(3.31)	-0.008	(-0.52)	-0.101	(-7.36)
Year Forecast Produced (Reference class = 2010 or later)						
Years before 2010	-0.007	(-5.64)	0.0002	(0.27)	0.003	(2.36)

* t values for Forecast Volume are relative to 1 not 0.

The results in equation format following Equation 4:

The set of models to estimate uncertainty around traffic forecasts based on our estimation results are:

$$
\begin{aligned}
y_{5,i} = {} & -182.26 + 0.705 * \hat{y}_i + 0.024 * \max(30{,}000 - \hat{y}_i, 0) + 0.006 * \text{Time Span} * \hat{y}_i \\
& - 0.006 * \text{County Unemployment Rate} * \hat{y}_i + 1 \\
& * (1 \text{ if Project is on a Freeway, 0 otherwise}) * \hat{y}_i - 0.15 \\
& * (1 \text{ if Project is on an Arterial, 0 otherwise}) * \hat{y}_i - 0.212 \\
& * (1 \text{ if Project is on a Collector, 0 otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if project is on an Existing Road, 0 otherwise}) * \hat{y}_i + 0.093 \\
& * (1 \text{ if project is on a New Road, 0 otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if forecast is done using Trend Analysis, 0 otherwise}) * \hat{y}_i + 0.068 \\
& * (1 \text{ if forecast is done using a Travel Model, 0 otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if forecast is produced after 2010, 0 otherwise}) * \hat{y}_i - 0.007 \\
& * (1 \text{ if forecast is produced before 2010, 0 otherwise}) * \hat{y}_i + \varepsilon_{P,i}
\end{aligned}
$$

$$
\begin{aligned}
y_{50,i} = {} & 255.55 + 0.891 * \hat{y}_i - 0.004 * \max(30{,}000 - \hat{y}_i, 0) + 0.008 * \text{Time Span} * \hat{y}_i \\
& + 0.002 * \text{County Unemployment Rate} * \hat{y}_i + 1 \\
& * (1 \text{ if Project is on a Freeway, 0 otherwise}) * \hat{y}_i - 0.062 \\
& * (1 \text{ if Project is on an Arterial, 0 otherwise}) * \hat{y}_i - 0.126 \\
& * (1 \text{ if Project is on a Collector, 0 otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if project is on an Existing Road, 0 otherwise}) * \hat{y}_i - 0.008 \\
& * (1 \text{ if project is on a New Road, 0 otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if forecast is done using Trend Analysis, 0 otherwise}) * \hat{y}_i - 0.008 \\
& * (1 \text{ if forecast is done using a Travel Model, 0 otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if forecast is produced after 2010, 0 otherwise}) * \hat{y}_i + 0.0002 \\
& * (1 \text{ if forecast is produced before 2010, 0 otherwise}) * \hat{y}_i + \varepsilon_{P,i}
\end{aligned}
$$

$$
\begin{aligned}
y_{95,i} = {} & 976.78 + 1.254 * \hat{y}_i - 0.413 * \max(30{,}000 - \hat{y}_i, 0) + 0.02 * \text{TimeSpan} * \hat{y}_i \\
& + 0.01 * \text{Unemployment Rate} * \hat{y}_i + 1 \\
& * (1 \text{ if Project is on a Freeway, } 0 \text{ otherwise}) * \hat{y}_i - 0.116 \\
& * (1 \text{ if Project is on an Arterial, } 0 \text{ otherwise}) * \hat{y}_i - 0.321 \\
& * (1 \text{ if Project is on a Collector, } 0 \text{ otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if project is on an Existing Road, } 0 \text{ otherwise}) * \hat{y}_i - 0.09 \\
& * (1 \text{ if project is on a New Road, } 0 \text{ otherwise}) * \hat{y}_i + \;\; 1 \\
& * (1 \text{ if forecast is done using Trend Analysis, } 0 \text{ otherwise}) * \hat{y}_i - 0.101 \\
& * (1 \text{ if forecast is done using a Travel Model, } 0 \text{ otherwise}) * \hat{y}_i + 1 \\
& * (1 \text{ if forecast is produced after 2010, } 0 \text{ otherwise}) * \hat{y}_i + 0.003 \\
& * (1 \text{ if forecast is produced before 2010, } 0 \text{ otherwise}) * \hat{y}_i + \varepsilon_{P,i}
\end{aligned}
$$

Overall Distribution: The overall distribution includes the intercept (α) and the forecast volume (β). We can think of these values as a reference line, with the remaining terms (γ) in the model changing the slope of that reference line. By including the variable on volumes in excess of 30,000 ADT, we allow for the slope of that reference line to change. The estimated coefficients suggest that for high volume roads, the counted traffic may fall short of the forecast but is unlikely to exceed it.

Functional Class: Relative to freeways, the coefficients for arterials and collectors are negative across all percentiles, shifting the range of expected outcomes down. This indicates greater deviation of actual traffic from forecasts on arterials and collectors than on freeways. This deviation may occur due to technical limitations of the forecasting method. For example, forecast volumes on collectors and arterials are likely to be more sensitive to the details of road network coding and zone size than freeway volumes because traffic does not load directly onto freeways.

Project Type: Forecasts for new roads have a narrower range of expected outcomes than forecasts for existing roads. We find this result counter-intuitive because we might expect it to be more difficult to forecast traffic on a new road. Forecasters might recognize this challenge and approach the task with more care.

Forecast Method: Forecasts made using travel demand models have a narrower range of expected outcomes than forecasts made using traffic count trends, population growth rates or professional judgment. We assume travel models are more accurate because they better capture the underlying factors that drive traffic changes.

Year Forecast Produced: In this model, we considered the year in which the forecast was produced as a continuous variable, defined as the number of years before 2010. We find that older forecasts are less accurate than newer forecasts. This result could reflect improved forecasting methods and data, but it could also be due to the nature of the projects themselves. For example, among the set of all forecasts produced since 2010, smaller projects are more likely to be complete, and therefore in our data set, than large projects.

Time Span: The time span measures the years between when a forecast is made and the opening-year. Time span coefficients are positive and significant for all percentiles, meaning that the post-opening traffic volume is more likely to exceed a long-term forecast than a short-term forecast. This result may reflect overall traffic growth rates exceeding expectations and slowly building over time. However, longer-term forecasts are likely to be associated with larger projects, so distinguishing the potential effects of project size and time span remains difficult.

Unemployment Rate: The coefficients are not significant at a 95% confidence level, but they suggest that forecasts made at times of high unemployment are less accurate. Predicting economic growth in the wake of a recession may be especially uncertain.

We tested a number of other model specifications that are not included in the final model. For example, we found no significant difference between urban and rural counties, and we tested using categorical variables for groups of years instead of a continuous variable for years before 2010. We tested interaction effects between area type and functional class, and between agency and forecast methodology and found that they did not improve the model.

6.4.3 Model Application

Given a forecast, we can apply the full model to estimate a range of expected outcomes, as the four examples in Figure 34 illustrate. The horizontal axes indicate forecast ADT, while the vertical axes show the range of expected outcomes. The perfect forecast lines in the figures correspond to traffic volume equal to the forecast. Example A shows a forecast made in 2019, using a travel model, for an existing arterial when the state unemployment rate was 4%. For a forecast of 30,000 ADT in year 2024 (a time span of 5 years), the chart indicates that we expect 90% of future traffic volumes to fall in the range of 19,000 to 36,000, with a median of 26,000. Consider a different project in example B that has the same parameters but a forecast of 30,000 ADT in year 2029 (a time span of 10 years). The uncertainty window for Example B is slightly wider with an expected range of 20,000 to 39,000 and a median of 27,000. Example C matches B but assumes a traffic count trend instead of a travel model, widening the uncertainty window to 21,000 to 46,000 with a median of 32,000. Example D assumes a collector instead of a minor arterial, which shifts the uncertainty window downward to 16,000 to 36,000 with a median of 26,000. In each of these examples, the median expected value is lower than the forecast.

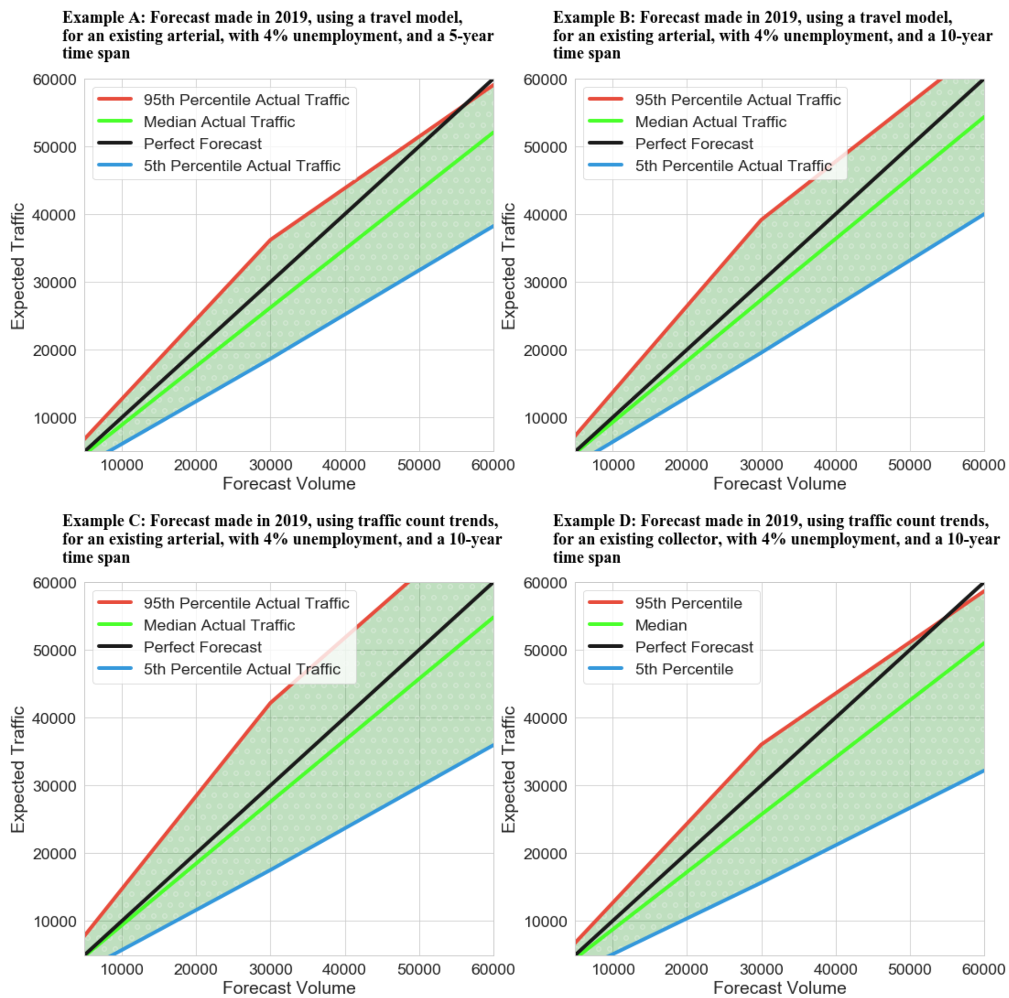

Figure 34: Expected range of traffic as a function of forecast traffic for four full model examples

6.5 Transit Ridership Forecast Uncertainty Model

Following the econometric framework to determine bias and quantify uncertainty presented in Hoque et al. (Hoque, Erhardt, Schmitt, Chen, and Wachs 2021), we present in this section two quantile regression models. We estimated several specifications using both linear and logarithmic model and present here the ones with the best fit. The first model, referred to as the Base Model, indicates the overall uncertainty in transit ridership forecasts. The second one, termed as the full model, introduces additional variables to incorporate reference-class forecasting in constructing the uncertainty window.

6.5.1 Base Model

We present the results from the regression analysis of the log of observed ridership against the log of forecast ridership in Table 21. Here the statistically significant results are highlighted in gray. The results clearly indicate the presence of statistically significant optimism bias in the transit ridership forecast for all percentile values. The median observed ridership for example is about 50% of the forecast ridership. We plot the observed ridership against forecast ridership in Figure 33. The perfect forecast line represents observation equal to forecast and is presented as a reference. We can see that above a forecast of 60,000 riders, 95% of the observed ridership is less than the forecast.

Table 21: Quantile regression results for base model of average weekday transit ridership

	5th Percentile		**50th Percentile**		**95th Percentile**	
Pseudo R-Squared	0.516		0.653		0.879	
	Coef.	(t value)	Coef.	(t value)	Coef.	(t value)
Overall Distribution						
Intercept (α)	-1363.49	-1.64	1709.25	2.03	7051.37	1.63
Forecast Ridership (β)*	0.335	13.77	0.502	4.517	0.808	6.19

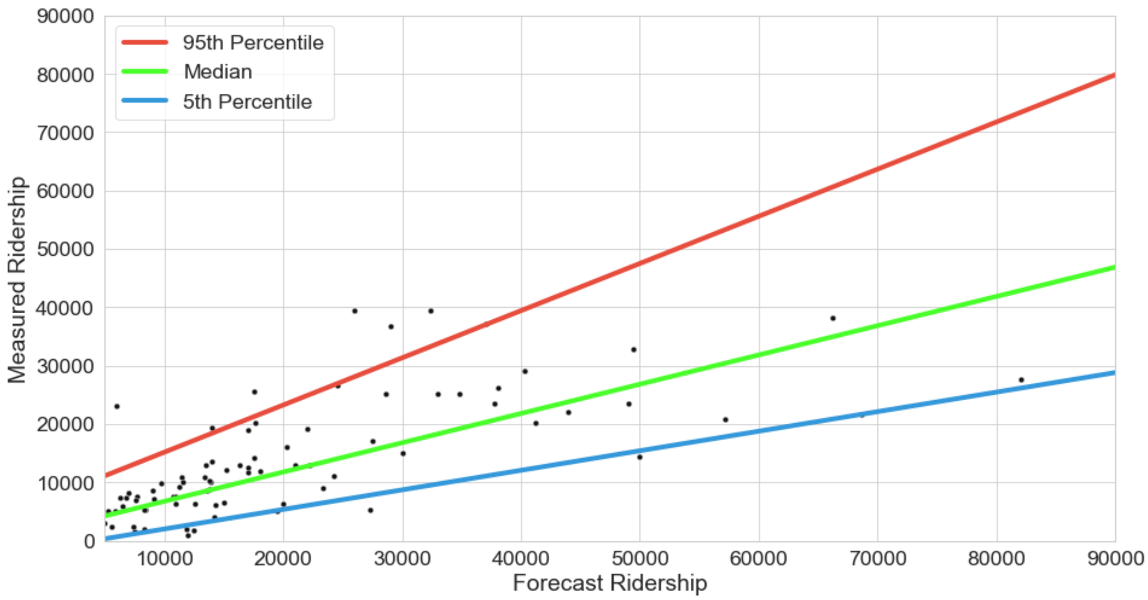

Figure 35: Observed versus forecast transit ridership and base model quantile regression lines

The results in equation format following Equation 4:

$$y_{5,i} = -1363.49 + 0.335 * \hat{y}_i + \varepsilon_{P,i}$$

$$y_{50,i} = 1709.25 + 0.502 * \hat{y}_i + \varepsilon_{P,i}$$

$$y_{95,i} = 7051.37 + 0.808 * \hat{y}_i + \varepsilon_{P,i}$$

6.5.2 Full Model

Results of the quantile regression model of the observed ridership against forecast ridership and other variables are presented in Table 22. As explained in the Methodology section, we considered the multiplicative effect of each variable except the intercept and the forecast ridership in our model so that their effects materialize in the changing slope of the quantile regression lines, i.e., they interact with the forecast itself by scaling it up or down. The intercept and the coefficient on forecast ridership serve as reference lines which change slope as each descriptive variable interacts with the forecast to affect the observed ridership.

For predicting the uncertainty envelope, we modeled the tails of the distribution. We tested several specifications with explanatory variables internal (e.g., transit mode, year of forecast, time span etc.) and external to the forecast (e.g., socio-demographic characteristics of the service area). The limitation hindering such models were the absence of data on key variables, in addition to a small sample size. Again, the explanatory variables apply to subsets of the sample, making statistically significant analysis difficult. Moreover, the variables themselves interact among themselves in a way that they sometimes change the direction of the coefficients and make the model unexplainable. We therefore estimated two separate models for the 5th and 95th percentile values with the 50th or conditional-median regression line as reference. The full model presented below contains the variables that make sense and are known at the time of forecast production in addition to producing better goodness-of-fit measure and consistent coefficients across the descriptive variables. The greyed cells represent the variables that are not significant at a 90% confidence level but kept in the final model because they produce a better fit. The interaction of these variables has been tested, but the interpretation is not clear. We discuss the implication of each variable below:

Table 22 Quantile regression results for full model of average weekday ridership

	5th Percentile		50th Percentile		95th Percentile	
Pseudo R-Squared	**0.49**		**0.84**		**0.96**	
	Coef.	**t value**	**Coef.**	**t value**	**Coef.**	**t value**
Overall Distribution						
Intercept (α)	-243.78	-0.57	81.43	0.15	89.28	0.11
Forecast Ridership (β)	0.005	0.03	0.66	4.44	1.08	7.61
Forecast Greater than 20k	0.28	1.49	-0.59	-2.58	-0.94	-4.60
Project Opening After 2005 (Before 2005 as reference)						
Number of years after 2005	0.06	2.52	0.04	1.96	0.05	1.74
Number of years after 2012	-0.12	-2.25	-0.02	-0.38	-0.07	-0.99
Project Service Area						
Project in the Central Business District	0.12	1.55	0.24	2.77	0.41	4.66
Mode (Reference: Urban Light and Heavy Rail Transit)						
Bus			-0.39	-2.98	-0.51	-2.43
Bus Rapid Transit			-0.51	-2.90	-0.27	-1.28
Commuter Rail			-0.37	-2.93	-0.45	-3.83
Streetcar			-0.70	-2.00	0.18	0.34
Ramp Up Period						
Number of years after project opening			0.03	0.44	0.001	0.02
Time Span						
Number of years between start and opening			0.003	0.16	-0.01	-0.59

The results in equation form are:

$$y_{5,i} = -243.8 + 0.005 * \hat{y}_i + 0.28$$
$$* \max(0, \hat{y}_i) + 0.06$$
$$* Number\ of\ Years\ after\ 2005\ the\ Project\ Opened * \hat{y}_i$$
$$- 0.12 * Number\ of\ Years\ after\ 2012\ the\ Project\ Opened$$
$$* \hat{y}_i + 0.12 * (1\ if\ Project\ Serves\ the\ CBD, 0\ otherwise)$$
$$* \hat{y}_i + \varepsilon_{p,i}$$

$$y_{50,i} = 81.43 + 0.66 * \hat{y}_i - 0.59 * \max(0, \hat{y}_i - 20{,}000) + 0.04$$
$$* Number\ of\ Years\ after\ 2005\ Project\ Opened * \hat{y}_i$$
$$- 0.02 * \text{Number of Years after 2012}\ Project\ Opened * \hat{y}_i$$
$$+ 0.24 * (\text{1 if Project Serves the CBD, 0 otherwise}) * \hat{y}_i$$
$$+ \quad 0 * (\text{1 if forecast is for an Urban Light or Heavy Rail})$$
$$- 0.39 * (\text{1 if forecast is for a Bus project}) * \hat{y}_i$$
$$- 0.51 * (\text{1 if forecast is for a Bus Rapid Transit project}) * \hat{y}_i$$
$$- 0.37 * (\text{1 if forecast is for a Commuter Rail project}) * \hat{y}_i$$
$$- 0.39 * (\text{1 if forecast is for a Streetcar project}) * \hat{y}_i$$
$$+ 0.03 * \text{Ramp Up Period} * \hat{y}_i + 0.003 * \text{Time Span} * \hat{y}_i$$
$$+ \varepsilon_{P,i}$$

$$
\begin{aligned}
y_{95,i} = \ & 89.28 + 1.08 * \hat{y}_i - \ 0.94 * \max(0, \hat{y}_i - 20{,}000) \\
& + 0.05 * Number\ of\ Years\ after\ 2005\ Project\ Opened * \hat{y}_i \\
& - 0.07 * \text{Number of Years after 2012}\ Project\ Opened * \ \hat{y}_i \\
& + \ 0.41 * (\text{1 if Project Serves the CBD, 0 otherwise}) * \hat{y}_i \\
& + \ 0 * (\text{1 if forecast is for an Urban Light or Heavy Rail}) \\
& - 0.51 * (\text{1 if forecast is for a Bus project}) * \hat{y}_i \\
& - 0.27 * (\text{1 if forecast is for a Bus Rapid Transit project}) * \hat{y}_i \\
& - \ 0.45 * (\text{1 if forecast is for a Commuter Rail project}) * \hat{y}_i \\
& + \ \ 0.18 * (\text{1 if forecast is for a Streetcar project}) * \hat{y}_i \\
& + \ 0.001 * \text{Ramp Up Period} * \hat{y}_i + \ 0.01 * \text{Time Span} * \hat{y}_i + \varepsilon_{P,i}
\end{aligned}
$$

Overall Distribution

The overall distribution includes the intercept (α), the forecast volume (β). We can think of these values as a reference line, with the remaining terms (γ) in the model changing the slope of that reference line. Average weekday ridership is lower than the forecast for the 5th and the 50th percentile values, but higher for the 95th percentile. However, there is a caveat introduced by the project scope. Transit projects differ in characteristics by their scope of service; variables like length, ridership, number of stops served by the project can be a substitute for the scope. In this model we used 20,000 weekday riders as a breakpoint and estimated a separate coefficient to account for larger projects. The positive and negative coefficients on 5th and 95th percentile signifies the narrow uncertainty window as we go to higher ridership forecasts. This needs to be kept in mind that this narrow forecast window may be the realization of lower percent deviation owing to a higher forecast, instead of forecasts getting better for high ridership corridors.

Project Opening Year

As we have surmised from our categorical exploration, forecasts have been getting better over the years. Each year after 2005, actual ridership got higher which led to lower percent deviation from forecasts. However, this positive effect is neutralized by the recent transit ridership trend, particularly that after 2012 as we have established in Chapter 4. We can infer from the number of years after 2012 variable that each year after 2012 saw lower average ridership and higher deviation from forecasts. These effects are not statistically significant for higher quantiles, but they point to forecasts getting better over the years and external factors influencing forecast performance.

Mode

Urban Light and Heavy Rail transit has observed a steady growth since 1990, which has resulted in lower average deviation from forecasts. All else remaining the same, the uncertainty window for bus and commuter rails are narrower than urban rail while that for bus rapid transit and streetcars are wider. There is no discernible effect of the modes on the 5^{th} percentile values, meaning the lower quantiles are independent of project mode.

Projects serving the Central Business District

Projects that serve the CBD of a metro area sees higher ridership than forecasts for all quantiles. It is generally assumed that work travel patterns are easier to model than non-work travel because of the publicly available home-to-work records in the American Community Survey and Longitudinal Employee-Household Dynamic (LEHD) data. Since the travel pattern in the CBDs are relatively easier to predict, it is associated with greater accuracy.

Ramp Up Period

Each year after project opening, ridership experiences a growth and therefore the deviation from forecasts get smaller, signified by the positive coefficient. The ramp up period considered in our dataset is less than or equal to 2 years; the smaller variation may have resulted in a statistically insignificant coefficient. Even so this variable is included in the model since the effect is well documented in practice and may give practitioners a guideline for selecting the observation year.

Time Span

The time span measures the years between when a forecast is made and the opening-year. Time span coefficients are positive for the 50th and negative for the 95th percentile, meaning that for the upper 50%, the uncertainty window is narrower. The 5th quantile values are not affected by this variable.

6.5.3 Model Application

The model results can be applied in an equation to construct an uncertainty window around forecast as demonstrated in two examples. We consider two nearly identical projects that opened in 2019 that serves the CBD of a metro area. In the first example (Figure 36), we consider a Light Rail project, and in the second (Figure 37), we consider a Bus Rapid Transit project. The shaded region in the figures represent the 90% uncertainty window for observation two years after project opening. If the forecast average weekday ridership were 15,000, our models estimate an uncertainty window of 1,631 to 26,171 for the LRT and 1,630 to 22,065 for the BRT project, the values being estimated by the 5th and the 95th percentile models.

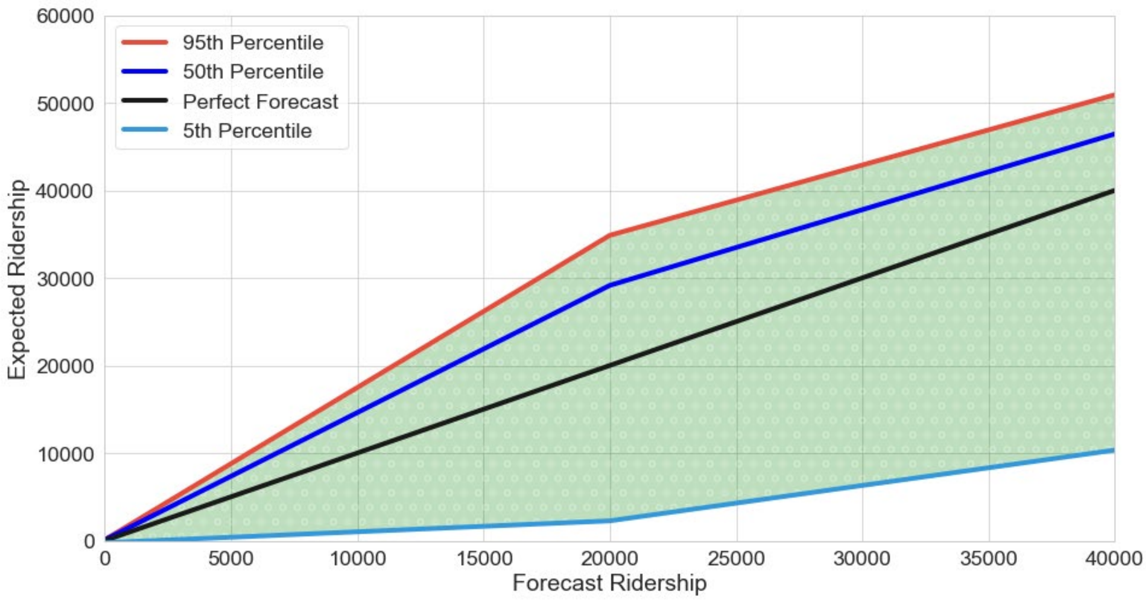

Figure 36: Expected vs Forecast Ridership Example A

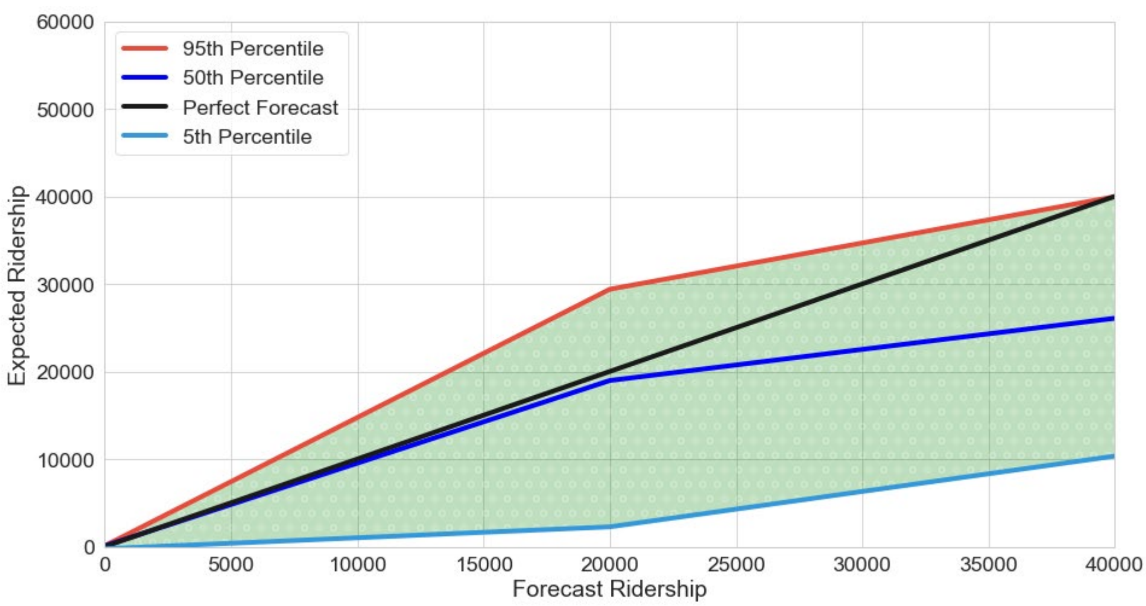

Figure 37 Expected vs Forecast Ridership Example B

6.6 Summary of Findings

This paper demonstrates how to use empirical measures of the accuracy of past traffic forecasts to estimate the uncertainty expected of future forecasts. In this discussion, we acknowledge the limitations of the work and offer recommendations for how planners should use it to improve traffic forecasting practice.

6.6.1 Limitations and Future Work

Traditional methods for estimating uncertainty in traffic forecasting rely on assumed ranges of inputs. Our method relies on data that include the full set of deviations occurring in the past, including the travel effects of events such as the 2008 financial crisis and fluctuating gas prices. However, this data-driven approach may be limiting if the future looks discontinuous from the past. For example, the effect of self-driving vehicles may pose a risk to forecasts made for 2040, and outcomes for projects that have already opened cannot clarify that risk. The National Road Traffic Forecasts (NRTF) in the UK recognizes this challenge and addresses them by investigating factors that most influence road traffic and their relation to such unknown and imminent changes in travel behavior (introduction of connected and autonomous vehicles in the network, changes in transportation policy etc.) (Lyons and Marsden 2019).

We estimated the quantile regression models in this paper using data that we assembled based on availability. It is the largest known data set of traffic forecast accuracy, but the data are not necessarily representative of transportation projects in general and are limited to data from a handful of agencies. We expect a reference class of projects similar to the project in question to provide the most reliable uncertainty estimates. That reference class could be a subset of these data, or it could be locally collected data on forecast and observed traffic volumes which the forecaster uses to estimate new quantile regression equations.

This study evaluates the uncertainty of Average Daily Traffic forecasts for road projects, which is only one type of transportation forecast. Future work could extend this approach to apply to other forecast variables, such as travel times or peak-hour traffic, or to other types of projects, such as transit ridership forecasts.

6.6.2 Recommendations for Practice

In spite of past calls to better consider uncertainty, single-point traffic forecasts remain the norm for most applications. We reiterate the call for transportation agencies to acknowledging uncertainty as an element of all forecasting, and recommend they do so by adopting three practices:

- Use a range of forecasts to communicate uncertainty. A forecaster can apply the quantile regression equations reported here to calculate the expected (median) traffic volume from a forecast, and the range within which to expect 90% of post-opening traffic volumes to fall. These equations are a function of the forecast volume and other project attributes, they require only a small additional effort to apply after creating a forecast. While other methods of estimating uncertainty, such as scenario testing and Monte Carlo simulation, may also be appropriate, the method presented provides the advantage of an outsider's view.
- Apply decision intervals to determine whether a forecast at the high or low end of the range would change an investment decision. Anam, Miller, and Amanin (2020) offer an approach for managing forecasting risk using decision intervals that identify the breakpoints at which a project decision would change. If a traffic volume at the low or high end of the uncertainty window would not change the decision, then planners can safely proceed with little worry about the risk of an

inaccurate forecast. Conversely, if the decision would change with a traffic volume at the extremities of the range, planners might seek to better understand the risks involved, or may choose an alternative with lower risk.

- Systematically monitor traffic forecast accuracy and use the resulting data to better estimate uncertainty. We estimated the equations in this paper from data shared by several transportation agencies in the U.S. and Europe. Other agencies involved in forecasting may use different methods, forecast for different types of projects, or be subject to different external conditions. Those agencies should collect local data tracking the accuracy of their own forecasts and use those data to estimate quantile regression models specific to their own situations. Because it is more difficult to assemble the necessary data after the fact, we recommend that agencies archive their forecasts at the time they are made, then add measured traffic outcomes after the project opens. NCHRP Report 934 provides recommendations on establishing such a data collection program, including the specific data items to record and how to archive the data efficiently.

By implementing these recommendations, agencies can better manage the risk inherent in forecasting. At times this may mean choosing a lower-risk alternative, and at times it may mean accepting the risk of a preferred alternative. Implementing these recommendations also allows agencies to better protect their credibility as forecasters. Whereas a point-forecast 15% different from the post-opening count might be viewed as inaccurate, the same forecast may be viewed as accurate if it were reported with a range of +/- 20%. By monitoring the accuracy of their forecasts, agencies document their track record, and can demonstrate that their uncertainty estimates are grounded in data on historical accuracy.

Chapter 7 CONCLUSION

This research investigates traffic and transit ridership forecasts from two largest databases of their kind to establish the evidence of uncertainty in forecasts. It explores the biases introduced by different project and forecast characteristics and establishes a relation between accuracy and uncertainty. Finally, this research presents a new tool, Quantile Regression Method, to quantify the accuracy in forecasts using past accuracy.

The overall conclusions from this work is presented in this chapter. I review the specific findings of each individual element that make up the research and put it in the context of the overarching research objectives. Section 7.2 proposes directions this study can be expanded into as a guide to future endeavors. Finally, I discuss the broader implications of this research for the field of travel demand modeling.

7.1 Research Findings

The research began with a question: how can we make forecasts that are good enough for policy decisions that hinges on huge investments? In the context of the Bipartisan Infrastructure Law of nearly $550 billion investment in transportation infrastructure in highway and public transit programs, the implications of this question are huge. Since investments of public dollars are informed by anticipated demand, ensuring these forecasts are good enough to inform the policy decisions is critical for accurate benefits to cost estimates.

One definition of "good enough" is that the forecast is close enough to the actual outcomes that the decision would remain the same if the decision had been made with perfect knowledge. For example, if the forecast is used to make a decision about how many lanes to build on a roadway, the conventional wisdom is that the traffic forecast should be “accurate to within half of a lane”. A corollary definition of “good enough” is that decision makers are willing to accept the consequences of a sub-optimal decision as a trade-off for the ability to move forward with imperfect information. If the consequences of an imperfect decision are low, then fewer resources can be invested in forecasting, whereas more extensive study and more accurate forecasts may be warranted when the consequences are high. This will naturally distinguish between smaller routine projects, and the larger mega-projects, or projects that are otherwise unique.

With the goal of aiding planners make informed policy decisions about future highway and public transportation projects, this research aims to quantify the uncertainty inherent in these decisions. The specific objectives set at the beginning of this study are:

- To establish empirical evidence of uncertainty in travel demand (traffic and transit ridership) forecasts
- To identify factors affecting the uncertainty in traffic and transit ridership forecasts and
- To develop quantile regression models to quantify the uncertainty in these forecasts.

In this dissertation, I have explored these objectives through four distinct elements presented in Chapters 3 to 6. In the following subsections, I will be discussing the results in the broad context of the research objectives.

7.1.1 Empirical Evidence of Travel Demand Forecast Uncertainty

The analysis of uncertainty in traffic and public transit ridership forecasts is based on two databases created as part of the project. In the Traffic Forecast Accuracy Database, we compiled about 2600 unique projects comprising of about 16000 segments. In the Transit Ridership Forecast Accuracy Database, we have 164 transit projects in the United States. While these two databases are the largest of their kind, not all the records were used in our analysis since a lot of them didn't open at the time of analysis and/or had multiple datapoints missing. The detailed criteria for selection of projects for analysis are described in Chapter 3 for Traffic Forecast Accuracy and in Chapter 5 for Transit Ridership Forecast Accuracy.

Conforming to the existing literature on forecast accuracy, our analysis point to significant optimism bias in travel demand forecasts. The measured traffic is on average 6% lower than forecast volume and ridership is about 24.6% lower than forecasts on average. The mean absolute difference between measured traffic volumes and forecasts was 17%. In addition, 90% of opening-year traffic volumes were in the range of -38% to +37% of the forecast volumes. This spread of outcomes persists after adjusting for the shift due to the Great Recession, suggesting that there are reasons for inaccuracy beyond this unforeseen event. For transit ridership forecasts, the mean absolute deviation is 40.2% with 90% of the observation falling between -81.6% and 45% of the forecast ridership. These values highlight that we should not consider travel demand forecasts to be point estimate, but a range of possible outcomes.

We found evidence of travel demand forecasts getting better over the years, but significant variability remains (Figure 38). Better data and forecasting techniques may have contributed to this improvement. But this can also be an effect of aggregate travel trends. Vehicle miles traveled (VMT) per capita grew rapidly in the 1980s and 1990s. In the 2000s, this trend leveled off and declined, before subsequently rebounding in about 2013. While VMT per capita was increasing, counted traffic volumes were higher than forecast, but after VMT per capita peaks, the opposite is true. This relationship suggests that traffic forecasts may not have fully captured the factors driving aggregate VMT trends, especially in the later years. Transit ridership also saw an unexpected decline since 2012 owing primarily to declining gas prices, economic growth, and emergence of new mobility options. If the forecasts are adjusted for these factors, the forecast performance of the hypothetical scenario would have seen further improvement.

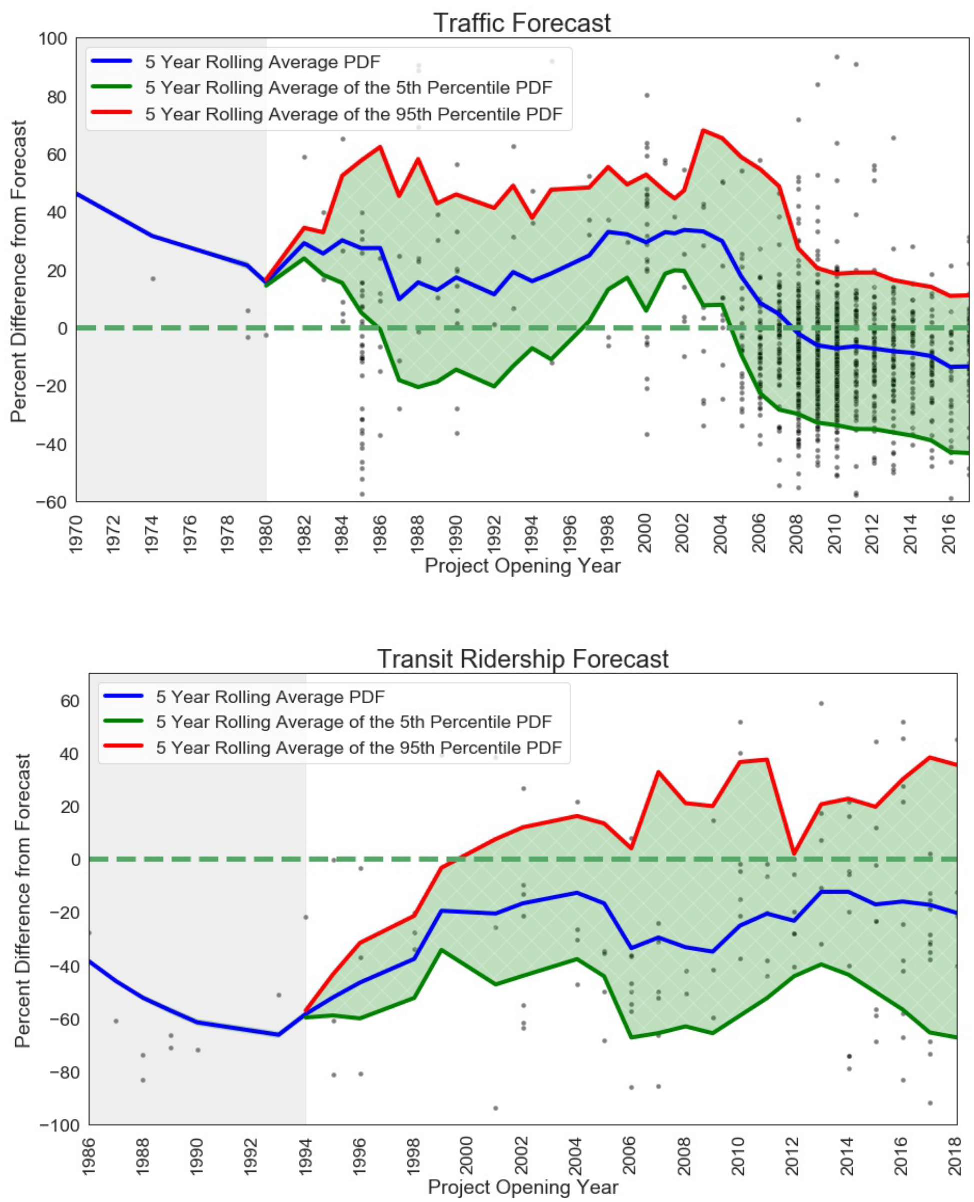

Figure 38: Variability in Forecast Performance over the Years

7.1.2 Factors affecting travel demand forecast uncertainty

Traffic and transit ridership forecasts have been improving over the years and this is more likely due to advances in forecasting techniques, availability of better data and models to understand travel behavior. In our investigation, we found several factors that have a bearing on forecast accuracy and uncertainty. Exogenous forecasts, unexpected changes in travel behavior due to economic shifts, project characteristics etc. affect forecast performance. We summarize our conclusions below:

- **Forecast method have an impact on forecast accuracy.** We found that travel models produced more accurate forecasts. Travel models are sensitive to the underlying determinants of traffic growth, including land-use changes and road network changes, so they were more accurate than traffic count trends. Methodological information for transit ridership forecasts is not available in our dataset to make such a comparison. However, the Capital Investment Grant (CIG) program by the FTA do present a chance to compare methodological advances in ridership forecasting. The Before-After studies as part this program has led to several advances in the industry: improved methods for forecast, application of risk assessment methodology and maintaining proactive oversight of project operation (Federal Transit Administration 2020). In 2001, FTA introduced new analytical tools which increased model scrutiny which may have resulted in better forecast performance. Ridership for projects that were produced after 2001 had mean PDF of -18.8% compared to -46% for the ones produced before this introduction.
- **Forecast Performance varies by time span**: We defined the time span as the number of years between the start year and the year of count. Traffic and transit ridership forecasts with a span of 5+ years were less accurate, and counts were lower on average than forecasts. On average, traffic deviated by about 23.7% and transit ridership by 44.8% from forecast for projects with time span greater than 5 years. The greater the number of years between forecast production and

measurement, the larger the opportunity for changes to have occurred in the economy, land use patterns, fuel prices, and other factors that influence travel. These are all variables that are difficult to predict, but their effects are evident.

- **Transit ridership forecasts vary by their location and area transit characteristics**. We found ridership for projects that are in the Central Business District (CBD) of the metro area has greater over-prediction than those outside. One possible explanation can be the dependence of ridership forecasts on employment, which is typically reflected in the CBD. Another important factor of transit ridership is the area's familiarity with transit systems. Project sponsors serving larger populations may have greater resources to devote to preparing rigorous forecasts. They may also answer to a wider variety of stakeholders, which could influence the incentives for promoting a particular project through optimistic forecasts. We tested this effect have considering the yearly operating expense of the transit agency. Our results show that forecasts in the metro areas with operating expense between \$100 million to \$300 million have in general better performing forecasts than those with operating expense greater than \$300 millions (-15.54% against -33.76%).
- **Factors affecting transit ridership affect their forecasts as well.** Transit ridership is dependent on the unemployment rate, percent of zero-vehicle households and average gas price of the area as well. Prior research works have established that growth of such economic factors is typically associated with higher transit ridership. In our analysis we found that the metropolitan statistical areas experiencing growing unemployment rates, zero vehicle households and

increasing gas prices from the start year to the observation saw lower average deviation from forecasts. As transit ridership grew in these metro areas, the deviation from forecasts got low as well.

7.1.2.1 Impact of the Recent Decline in Transit Ridership on Ridership Forecasts

We identified a number of factors that affect transit ridership, some of which result in increases and others in decreases to transit ridership. Together these factors result in a net bus ridership decline of 15% and a net rail ridership decline of 3% between 2012 and 2018. While several factors contribute to lower transit ridership, including lower gas prices, higher fares, and changes to income, teleworking rates and car ownership, we show that ride-hailing is the most important. By 2018, ride-hailing reduced bus ridership by 10% and reduced rail ridership in mid-sized metropolitan areas by a similar amount. It had a positive, but insignificant effect on rail ridership in the largest metropolitan areas.

- Transit connects people to activities and jobs, so the number and location of both affect transit ridership. Each 1% increase in population plus employment is associated with 0.22% more transit ridership. Similarly, higher density leads to more transit ridership. We considered the percent of the population and employment in a region that is within a transit supportive density, defined as more than 10 people or employees per acre. For each percentage point increase (such as from 10% to 11%) in population plus employment living in these denser areas, transit ridership is 0.4% higher.

- Higher gas prices make driving more expensive and incentivize travelers to use transit. Each percent increase in gas price accounts for a 0.14% increase in transit ridership.
- Several factors related to the characteristics of households, their income, and their work norms may affect transit ridership. We find three to be important: income level, 0-vehicle households, and telecommuting. With higher per capita income, people are less likely to ride transit. Each 1% increase in the median per capita income results in a 0.07% decrease in transit ridership.
- Higher shares of 0-vehicle households in an MSA have a small positive effect on transit ridership. We know that people from households without a car constitute an important market of transit riders. However, our estimated coefficient is small, so the results show that the change in vehicle ownership explains little about the change in transit ridership over this period. Our results show that a decrease from 10% of households owning 0 vehicles to 9% of households owning zero vehicles would result in 0.2% less transit ridership, but this effect is not statistically significant.

As unemployment, percent of zero-vehicle households and average gas prices increase in a metro area, people tend to use transit more. For optimistically biased forecasts, this means that actual ridership would be closer to the forecast which is what we see in our analysis. Ridership on projects that opened after 2012 was about 18.7% lower than forecast on average. After adjusting the ridership for metro area population, employment, household income, zero vehicle households and presence of TNCs, this PDF comes down to -12.2% (Figure 39). The absolute deviation decreases as well.

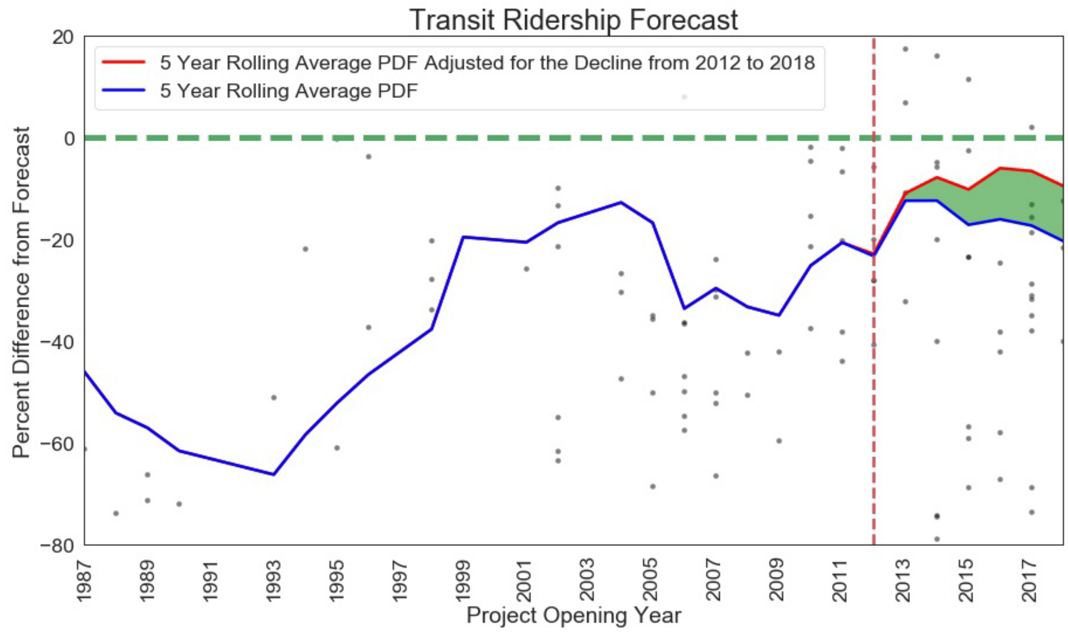

Figure 39: Transit Ridership Forecast Performance after adjusting for the decline since 2012

7.1.3 Quantile Regression Models to Convey forecast uncertainty

We estimated several quantile regression models to quantify the uncertainty window around a travel demand forecast. The model takes the following form:

$$y_{P,i} = \alpha_P + \beta_P \hat{y}_i + \delta_P X_i \hat{y}_i + \varepsilon_{P,i} \qquad \textit{Equation } 1$$

where y_i is the counted traffic on project i, $\hat{y}_i$ is the forecast traffic on project i and ε_i is a random error term. α and β are estimated regression coefficients, X_i is a vector of descriptive variables associated with project i, and δ is a vector of estimated model coefficients associated with those descriptive variables, and *P* index indicates that the term applies to the 5th, 50th or 95th percentile. The equation for the 95th quantile model, for example, would fit a regression line so that 95% of the datapoints fall below the line. This means that a combination of the 5th and 95th quantile lines would encase 90% of the observations and therefore can therefore predict the uncertainty based on past results.

7.1.3.1 Quantile Regression Model for Traffic Forecast Uncertainty

The final model presented in this research for estimating the uncertainty in traffic forecasts include several explanatory variables:

- **Functional Class of the roadway**: The coefficients indicate greater variation of actual traffic from forecasts on arterials and collectors than on freeways.
- **Project Type**: Forecasts for new roads have a narrower range of expected outcomes than forecasts for existing roads.
- **Forecast Method**: Forecast made with travel demand models are more accurate than trend-based results because they better capture the complexity in travel demand.
- **Year forecast produced**: Forecasts for newer projects, in particular those forecasted since 2010, were more accurate than older ones. This can indicate better forecasting technique, better data available. This can also point to the mix of projects since projects since 2010 are more likely to be small projects like resurfacing.
- **Time span**: Longer time horizon introduces additional uncertainties in land-use changes and induced demand. Forecasts get less accurate with each year after the year forecasts are produced.

Given a forecast, we can apply the full model to estimate a range of expected outcomes, as the example in Figure 40 illustrate. The horizontal axes indicate forecast ADT, while the vertical axes show the range of expected outcomes. The perfect forecast lines in the figures correspond to traffic volume equal to the forecast. Example A shows a forecast made in 2019, using a travel model, for an existing arterial when the state unemployment rate was 4%. For a forecast of 30,000 ADT in year 2024 (a time span of 5 years), the chart indicates that we expect 90% of future traffic volumes to fall in the range of 19,000 to 41,000, with a median of 27,000.

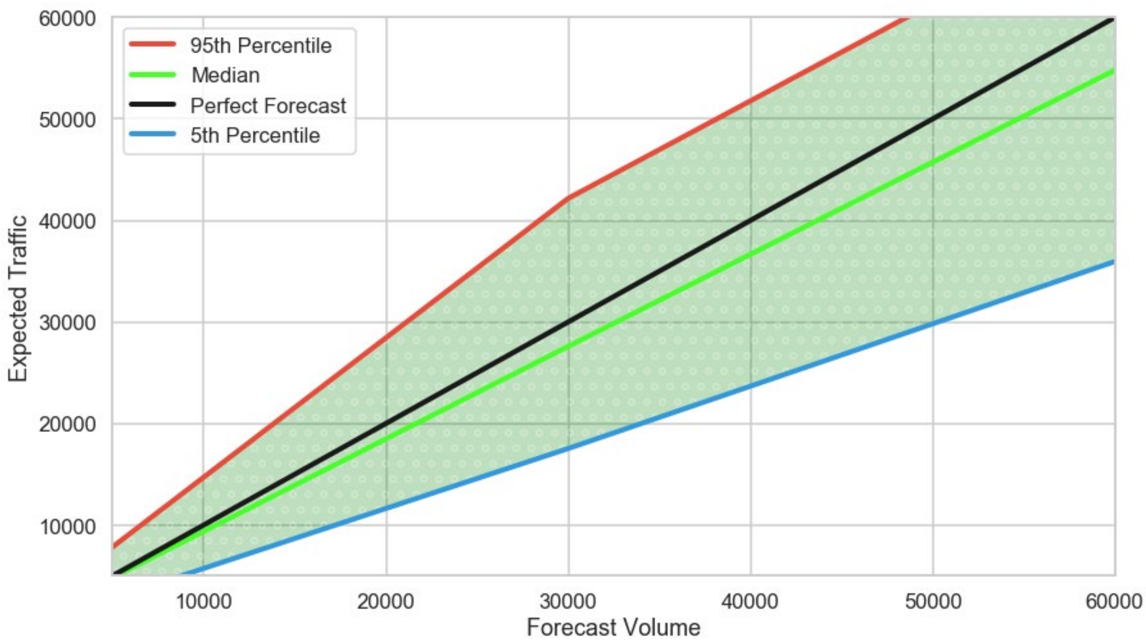

Figure 40: Uncertainty in a Traffic Forecast

7.1.3.2 Quantile Regression Model for Transit Ridership Forecast Uncertainty

The uncertainty in transit ridership forecasts, as we have found, depends on several factors and their effect is not uniform across the range. We have estimated three separate models to produce the uncertainty window:

- **Project mode**: The accuracy of transit ridership forecast varies by the mode. Taking Urban Light and Heavy Rail as reference, we see negative coefficients across the percentiles on other modes: signifying decrease in measured ridership in bus, bus rapid transit, commuter rail and streetcars.
- **Project location**: Projects serving the Central Business District have higher actual ridership. This means that ridership for these projects are closer to the forecasts.
- **Project opening year**: We used the year the project was opened for use as a substitute for the evolving forecasting methodology. Years after 2005 was associated with ridership closer to the forecast, signifying improvement in the methodology. It is to be noted here that the project opening year is used as a variable since year forecast was produced had a significant amount of data missing. Moreover, projects open several years after they have been forecasted for; most of the projects that opened on or after 2005 were forecasted after 2000, the year associated with updated methodology.
- **Project opening during declining ridership:** Projects that opened between 2012 and 2019 experienced lower ridership and therefore higher deviation from forecast. This decline is caused by several factors like metro area population and employment, zero vehicle households, emerging mobility in TNCs, bike and scooter shares, teleworking etc. Lack of data and limitations in sample size make most of these factors unquantifiable for ridership forecast uncertainty, however.

- **Project size**: Forecast greater than 20,000 average weekday riders have a net negative effect on the 50th and 95th percentile values. This can possibly be due to the large inaccuracies of the multi-line, heavy rail systems constructed during 1970-90.
- **Ramp Up Period**: Each year after opening is associated with higher ridership caused by the ramp up effect as demand matures after opening.
- **Time span**: Similar to traffic forecasts, longer time horizon introduces additional uncertainties in land-use changes and induced demand. Forecasts get less accurate with each year after the year forecasts are produced.

As an example of the application of the model results to construct an uncertainty window, let us consider a forecast for a Light Rail project serving the CBD of a metro area that opened in 2019. The uncertainty window is depicted by the green shaded region in Figure 41. For a forecast of 15,000 average weekday ridership for 2019, the model predicts a range of 1,600 to 26,171 actual ridership, these being the 5th and 95th percentile values respectively.

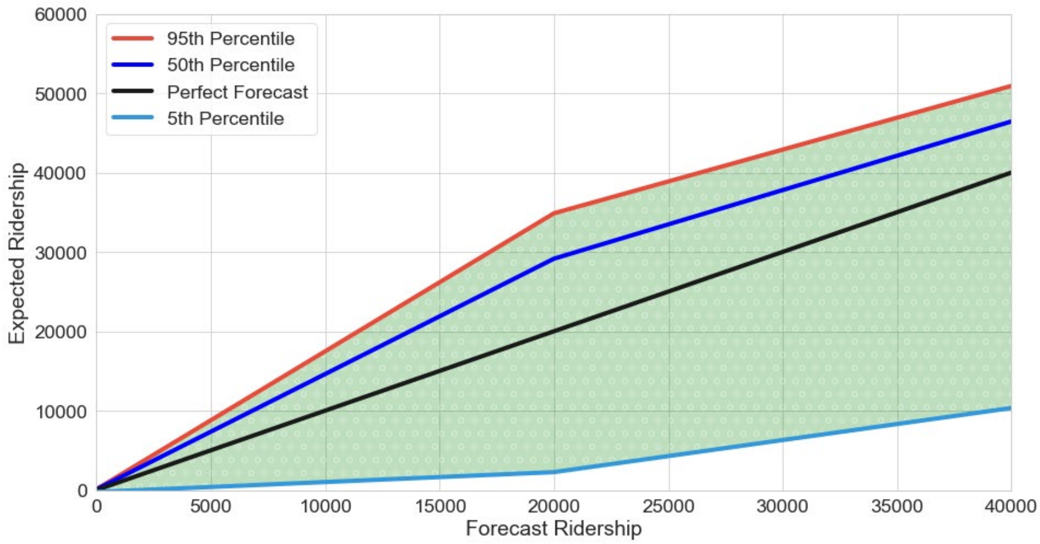

Figure 41: Uncertainty in a Transit Ridership Forecast

7.2 Future Research

This study bridges the gap of unknown forecast accuracy in the United States. It is by no means complete; as laid out in the previous section it is limited in the scope that it contains only a handful of state transportation agencies participating in the research. Availability of more data from states experiencing different economic growth than the one experienced by the participating agencies would make it easier to come to a more robust conclusion about the effect of unprecedented economic growth, positive or otherwise, can have on the accuracy of forecasts. In addition, the databases don't have complete information for most of the variables for comparison and model estimation. The participating state databases have, almost always, two-thirds of the data fields filled up, but it is never the same two-thirds. Depending on the availability of data about forecast methodology, relative accuracy of the different types of travel demand models, traditional 4 step models, activity- based models, and even different systems can be explored. This can be a measure of performance between the different models and help agencies identify the shortfalls of their own.

More broadly, the analysis of uncertainty presented in this study can be extended in several ways:

7.2.1 Effect of Uncertainty on Project Performance

The primary uses of travel demand forecasts are to justify the costs for a project and designing the same for a future scenario. Traffic forecasts, for example, are used to determine the average vehicle load on a roadway which in turn affects the thickness of the asphalt. Likewise, transit ridership forecasts determine how many buses will be in service at various times of the day. The ramifications for under-designing are great: a newly built roadway may need a resurfacing sooner than expected if traffic is much higher than expected, or a bus rapid transit may still fail to serve a significant portion of the target population. On the other hand, an over-designed project (addition of a lane when one less would have sufficed, or a new bus stop that isn't useful) may mean wastage of tax-payer money for very little benefit. It is therefore essential to quantify the impact of uncertainty in the overall project performance.

At this point we need to acknowledge the lack of standard for measuring project performance. The FTA employs the actual and predicted costs and ridership before-after as a metric for transit projects. A roadway project may be compared by the level of service, travel time savings, differences in average daily traffic and safety improvements.

The Highway England's Post-Opening Project Evaluations (POPE) framework may provide a useful direction in this case. The program reports several post-opening evaluations: traffic counts, costs, travel times etc. to verify the effectiveness of the project against its stated goals. Additional research in changes in peak period travel volumes, roadway speeds, truck volumes will continue to inform our understanding of the uncertainty around forecasts, as well as provide the opportunity to reduce that uncertainty as well as any bias present.

7.2.2 Uncertainty and Portfolio Performance

Uncertainty in travel demand forecasts have a direct effect on project performance in terms of design details and project goals. The effect of uncertainty goes beyond this, however, since they are used to choose between several alternatives. A potential extension of the study is the retrospective evaluation of alternatives along with their potential uncertainties based on the QR models presented here.

7.2.3 Effects of Travel Trend on Forecast Uncertainty

Our analysis show that the recent decline in aggregate demand for transit had a negative effect on the transit ridership forecast performance. The effect of the aggregate VMT trend on traffic forecasts are noticeable as well, even though it was not possible to quantify it. Future research can compare the VMT trends throughout the region to provide additional insights into whether any inaccuracy is specific to the project or regional in nature.

Moreover, the effect of the COVID-19 pandemic remains unexplored because of the recency. As COVID cases surged throughout, the world itself came to a gradual "lockdown", with traffic levels falling to an unprecedented level. By March 2022, it has almost recovered to the pre-pandemic levels but there might be sustained effects of it in the future. Percentage of people working from home has increased during this time, and may continue to stay at a much higher level than before. This can mean less traffic in the peak periods thereby increasing the difference from forecast. Future research needs to investigate the pandemic effect on travel demand forecasts.

7.3 Application in Transportation Planning

Forecasting the future demand for a facility, new or otherwise on a highway or a transit network, is a critical activity for project selection, design, and eventual measures of project success in terms of benefits to cost. Accurate forecasts for planning and design help ensure that public dollars are spent wisely. It is therefore in the interest of transportation planners and policy makers to base such decisions on the most accurate forecasts possible. However, as we have demonstrated in this research, forecasts are inherently uncertain. It is prudent to quantify the expected inaccuracy around traffic forecasts and consider that uncertainty in making decisions. Together, more accurate forecasts and a better understanding of the uncertainty around traffic forecasts can lead to a more efficient allocation of resources and build public confidence in the agencies that produce those forecasts. The contribution of this study in the broad context of transportation planning can be divided into four products:

- The data collected
- The evidence of accuracy and uncertainty presented
- The methodology established to communicate uncertainty in forecasts and
- The tool that uses quantile regression models to produce uncertainty windows around forecasts.

In the following subsections, I will be briefly presenting the lessons learned and contributions putting them in the framework of the four products.

7.3.1 Data collected for measuring accuracy

There is value in understanding the historic accuracy of forecasts in part because it provides an empirical means of communicating the uncertainty in outcomes surrounding a forecast. This ability is predicated on having the data to support such analyses. The absence of data has been the major barrier to the study of travel forecast accuracy (Nicolaisen and Driscoll 2014). This deficiency arose because accumulating the data needed for retrospective analysis requires proactive planning. The responsible agencies do not commonly preserve and archive forecasts, and so often lose these data. Long project development cycles and staff attrition make recovering this information cumbersome.

In this study, we assembled those data, compiling a database of forecast traffic for 2,611 unique road projects in six states in the United States (US) and four European countries. This resulted in the largest known database of forecast accuracy. The Transit Ridership Forecast Accuracy database used in this research is also the largest of its kind in the United States with 164 transit projects spanning from 1974 to 2019. Both of these databases contain project information (jurisdiction, area, highway functional class or transit mode, project type), forecast information (when it was produced, forecast traffic or ridership, methodology) and the observation information (year count was taken, units of measurement etc.). Our analysis, however, is based on subsets of these two databases, owing to several criteria we imposed for selection: presence of post-opening count, same year forecast and count data, and completeness of information. The data base is not a random sample of all highway projects, and this limits our ability to generalize from the analysis. The years in which projects in the database opened to traffic ranges from 1970 to 2017, with about 90% of the projects opening to traffic in 2003 or later. While the exact nature and scale of each project is not known in every case, almost half of the entries in the database were design forecasts for repaving projects. Earlier projects were more likely to be major infrastructure capital investment projects and more recent ones were more often routine resurfacing projects on existing roadways. This arose because some state agencies began tracking all forecasts as a matter of course only within the past ten to fifteen years and, in earlier years, information was retained only for major investments. In addition to the mix of projects in the database, there also were notable differences in the forecasting methods used across agencies. Because the traffic counts were of average daily traffic, comparisons could not be made of peak period traffic, by day of the week, or by season.

The database on transit ridership forecast is limited in the sample size for statistically significant analysis of factors affecting uncertainty. While it is still the largest database of its kind in the US, it doesn't have information in forecast methodology, assumptions in the demand model etc. making it difficult to compare across different forecasting techniques. The small sample also means that we weren't able to detect statistically significant effects of several variables that are reported to affect forecast performance as par literature.

Despite these limitations, the data collected for this study provides an important baseline for future analysis. The traffic forecast database, for example, has about 2611 unique projects out of which only 1291 were used in our analysis. Most of the rest didn't open at the time of analysis. As they open and the counts are added to the database, it provides the opportunity to expand the analysis. Data collected on actual project outcomes can be used as a benchmark against which to test a new travel model. Rather than focusing the validation only on the model's fit against base-year data, this would test whether the new model is able to replicate the change that occurs when a new project opens. This is akin to testing a model in the way it will be used, and a much more rigorous means of testing.

Furthermore, each of the records in our databases have multiple forecasts for different forecast horizons —opening year, mid-design year, design year etc.— available. Observations corresponding to these years can provide useful information about ramp-up effect, induced demand, and demand maturity, aside from quantifying the performance of travel demand models in long-term planning.

Another advantage of these two databases is providing forecasters with a library of projects similar to what they may currently be forecasting. This would enable what Flyvbjerg refers to as reference class forecasting (Flyvbjerg et al. 2006) and is especially valuable for new forecasters who do not have a lifetime of their own experience to draw from.

Updating and reporting forecast accuracy results with local data provides a better indication of the performance of the tools that a specific agency will use. This can document improvement or better than typical accuracy. If an agency has a track-record of accurate forecasts, using this data to update the quantile regression models will allow the ranges considered in Recommendation 1 to be narrower.

7.3.2 The Empirical Evidence of Uncertainty

Consistent with past research, our results show the distribution of actual traffic volumes and transit ridership around the forecast volume. These distributions provide a basic understanding of the uncertainty in outcomes surrounding a forecast. A goal of forecasting is to both to minimize the bias in this distribution, and to reduce the variance such that the forecasts more closely align with actual traffic. While our results show that forecasts have tended to improve over time, we cannot ever expect to achieve perfection in the realm of forecasting. In addition to a sub-optimal decision for a specific project, this inaccuracy may undermine the trust in forecasts made for other projects.

Getting rid of this inaccuracy by improving the modeling practice itself through better data, better models and better understanding of travel behavior is certainly the goal. However, it is evident from our analysis that there are always elements that introduce bias in the forecasts to varying degree by project, forecast and area characteristics. It is therefore prudent to be transparent about potential deviation from forecast and how the deviation would affect project success. The empirical evidence of uncertainty in travel demand forecasts as presented in this research presents the transparency necessitated.

While transparency does not necessarily ensure that forecasts will be accurate, it does send a clear message that the agency preparing the forecasts has nothing to hide, that any inaccuracies are the result of unexpected outcomes and not deliberate misrepresentation, and that the agency is legitimately interested in learning from those inaccuracies and using them to improve. If the agency can build a track record of accurate forecasts, it provides evidence with which to build trust in their abilities and establish the credibility of future forecasts. These benefits related to building credibility are in addition to the benefits associated with using the information to generate more accurate forecasts.

Our analyses reveal several factors related to accuracy. Among these are economic conditions, and we found evidence of a major unforeseen event—the Great Recession—causing a systematic shift in accuracy. In the wake of a much different disruption due to COVID-19, our results should open a discussion on communicating uncertainty in forecasting. It is reasonable to expect that there may be some major disruptive event within the scope of our next long-range forecasts. Moreover, such events are not the only factors contributing to forecast inaccuracy as a substantial spread of percent difference from forecasts remains after adjusting for the recession. Factors like forecast methodology, forecast horizon and project type affect the accuracy as demonstrated in this study, along with other unknown or unquantified factors.

For transit ridership forecasts, the unexpected decline in ridership from 2012 to 2018 had a substantial effect on their accuracy. Most of the decline can be attributed to the emerging mobility, especially the presence of ride-hailing services. In the next few years, several advances in the way people weigh travel options are expected. Emergence of Connected and Autonomous Vehicles (CAVs), wide-spread adoption of working from home as a (surprisingly positive) result of the pandemic, and market penetration and saturation of ride-hail, bike and scooter share services may affect both transit ridership and road traffic. In this scenario, the evidence of uncertainty presented in this research provides the impetus to recognize the limitations of our knowledge and set realistic expectations.

7.3.3 Methodology to Communicate Uncertainty in Forecasts

Reporting a range of forecasts explicitly communicates the risk associated with forecasts, and it is possible that the range results in a different decision, or the introduction of strategies to manage that risk. If the project decision would be the same across the range of forecasts, this adds confidence that the decision is defensible. How to do that is a point of contention in existing literature. A byproduct of the analysis presented in this research is the methodology presented in this research. Rather than relying on assumptions about inputs to the demand model, we demonstrate how to use empirical measures of past forecast accuracy to estimate the uncertainty in future forecasts. This constitutes what data to collect and archive, how to evaluate uncertainty and reference class forecasting. The process has been established as part of the NCHRP 08-110 Project: Traffic Forecast Accuracy Assessment Research (Erhardt et al. 2020) and verified by the Transit Ridership Forecast accuracy and uncertainty assessment section of this dissertation.

7.3.3.1 Data to be Collected and Archived

In our research, we found that we were able to learn more from projects where we had more information available. The basic project information available to the analysis allowed us to create the overall distributions of forecast accuracy, consider the effect of different factors, and generate the quantile regression models. The lesson learned from this exercise is that agencies need a standardized archival system to store project forecast information for periodically analyzing and reporting the accuracy of their forecasts. Archiving forecasts in a consistent manner reduces the time needed to analyze the forecast accuracy and strengthens any findings. A strong archival process ensures the necessary details about the forecasts are preserved in a readily accessible format once the project is opened to traffic. While some agencies are archiving some details of their project forecasts, the NCHRP 08-110 study revealed that archiving procedures were not consistently followed. For example, the traffic forecast accuracy study had to remove over 1,000 projects from the original collection of projects due to incomplete information. This amount reduced the project forecast database by nearly half. Strict archiving procedures would have greatly increased the study's database and strengthened its findings.

At the bare minimum, the database needs project, forecast and observation information. More detailed analysis is possible if it contains additional descriptive variables. The detailed data archival standards is presented in (G. Erhardt et al. 2020). We present a brief summary in Table 23.

Table 23: Database Summary

Category	Field	Traffic Forecast Accuracy Database	Transit Ridership Forecast Database
Project Information	Project Unique ID	✓	✓
	Transit Mode		✓
	Jurisdiction	✓	✓
	Location (City, County, Metro Area, State)	✓	✓
	Short Description	✓	✓
	Improvement Type (Widening, resurfacing, transit route extension, transit route redesign)	✓	✓
	Forecast year of completion	✓	
	Actual completion or opening year	✓	✓
	Length	✓	✓
	Competing/supporting modal system		✓
Forecast Information	Forecaster type (consultant, MPO)	✓	
	Forecast Traffic or Transit Ridership	✓	✓
	Forecast Year Type (Opening year, design year)	✓	

	Year forecast produced	✓	✓
	General methodology	✓	
	Post-processing applied	✓	
	Scenario information, if multiple forecasts are produced for the same project, e.g. build and no-build scenario.		
Observation Information	Observation (traffic count, transit ridership)	✓	✓
	Year of observation	✓	✓
	Units (AADT, Average Weekday Ridership)	✓	✓
Other information	Forecasting model details (4-step model, activity based model, multinomial logit etc.)		
	Vehicle Miles Travelled or Vehicle Revenue Miles (for both forecast and observation year)		
	Other descriptive variables for both forecast and observation year, e.g. local area population, employment, auto-ownership etc.		

7.3.3.2 How to Evaluate Uncertainty

We have established that accuracy and uncertainty are intertwined concepts. Since an uncertainty estimate is a "*means of expressing the accuracy of results*" (Collaboration for Nondestructive Testing n.d.), accuracy can act as an estimator of uncertainty. We defined the metric of accuracy as Percent Difference from Forecast (PDF):

$$PDF = \frac{Count - Forecast\ Volume}{Forecast\ Volume} * 100\%$$

Negative values indicate that the post-opening counted volume was lower than the forecast, and positive values indicate the counted volume was higher than the forecast. It expresses the deviation relative to the forecast, so provides meaningful information when making a forecast. We reported half the difference between the 5th and 95th percentiles as a measure of the spread of outcomes after adjusting for the average deviation. We separately reported the mean absolute PDF (MAPDF) as a measure of the general accuracy.

The count volumes to be compared to the forecasts should be taken on or after the forecast year and after opening if that information is available. Each year after opening, the project experiences ramp-up and induced demand effect, thereby skewing the accuracy results. In this research, we used the earliest post-opening count for traffic forecasts keeping consistent with existing studies. For transit ridership forecast accuracy, we considered a maximum of two year ramp up as par FTA practice (Federal Transit Administration 2020).

7.3.3.3 Reference Class Forecasting using Quantile Regression

Two methods are commonly adopted in practice to produce a range of forecasts instead of a point estimate— sensitivity tests and scenario analysis. Both of these approaches consider possible uncertain events or parameters and builds up to a range: an approach termed as the traditional insider's approach (Ascher 1979; Bent Flyvbjerg 2007b). On the other hand, it is possible to consider a project relative to a statistical distribution of past outcomes from a reference class of projects. For example, in project scheduling, the insider's approach estimates the duration of each task and sums to a total, whereas the outsider's approach looks at the average duration of similar completed projects. Flyvbjerg (2007) recommends the use of reference class forecasting for large infrastructure projects. In this study, we propose quantile regression as a tool for reference-class forecasting, seeing it models the tails of the conditional distribution. A 5^{th} percentile quantile regression model, for example, would fit a regression line through the data so that 95% of the observations are above the line. Two regression lines for 95^{th} and 5^{th} percentile would therefore produce a range a values within which 90% of the observation fall (Lê Cook and Manning 2013). The econometric framework for estimating such model is:

$$y_{P,i} = \alpha_P + \beta_P \hat{y}_i + \delta_P X_i \hat{y}_i + \varepsilon_{P,i}$$

Where y is the observation (traffic count or ridership), $\hat{y}$ is the forecast volume, X_i is a vector of descriptive variables associated with project i, and δ is a vector of estimated model coefficients associated with those descriptive variables, and P index indicates that the term applies to the different percentiles.

If an agency has collected data on forecast accuracy, the quantile regression models can be estimated using local data. Doing this is advantageous because it is based on data that are likely more similar to the types of forecasts that an agency will continue to perform. It is important that projects used to develop the quantile regression equations be (1) sufficient in quantity to produce statistically significant coefficient estimates and (2) representative of all the types of forecasts made. If an agency does not have a sufficient sample of local projects to support model estimation, it should supplement their local data with data from projects at peer agencies. The data provided with this report can be used. It is also recommended to use a census of all (not a sample) projects to the extent possible. This will avoid "cherry picking" highly accurate or inaccurate forecasts.

7.3.4 Quantile Regression Models

The quantile regression models of traffic and transit ridership presented in this research utilizes two largest and most complete databases of their kind. The models themselves are selected from different specifications to have the best fit, with maximum statistically significant explanatory variables for each quantile. In the absence of models estimated by an agency using their local data and incorporating internal complexity, these models provide the baseline for estimating uncertainty around forecasts. A forecaster can apply the quantile regression equations reported here to calculate the expected (median) traffic volume from a forecast, and the range within which to expect 90% of post-opening traffic volumes to fall. These equations are a function of the forecast volume and other project attributes, they require only a small additional effort to apply after creating a forecast. Anam, Miller, and Amanin (2020) offer an approach for managing forecasting risk using decision intervals that identify the breakpoints at which a project decision would change. If a traffic volume at the low or high end of the uncertainty window would not change the decision, then planners can safely proceed with little worry about the risk of an inaccurate forecast. Conversely, if the decision would change with a traffic volume at the extremities of the range, planners might seek to better understand the risks involved, or may choose an alternative with lower risk.

REFERENCES

Adler, Thomas, Michael Doherty, Jack Klodzinski, and Raymond Tillman. 2014. "Methods for Quantitative Risk Analysis for Travel Demand Model Forecasts." *Transportation Research Record* 2429 (1): 1–7. https://doi.org/10.3141/2429-01.

Aldrete, Rafael, Arturo Bujanda, Gabriel A Valdez-Ceniceros, and others. 2010. "Valuing Public Sector Risk Exposure in Transportation Public-Private Partnerships." University Transportation Centers Program (US).

Alemi, Farzad, Giovanni Circella, Susan Handy, and Patricia Mokhtarian. 2018. "What Influences Travelers to Use Uber? Exploring the Factors Affecting the Adoption of on-Demand Ride Services in California." *Travel Behaviour and Society* 13 (October): 88–104. https://doi.org/10.1016/j.tbs.2018.06.002.

Anam, Salwa, John S. Miller, and Jasmine W. Amanin. 2020. "Managing Traffic Forecast Uncertainty." *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering* 6 (2): 04020009. https://doi.org/10.1061/AJRUA6.0001051.

Anderson, Michael D, Walter C Vodrazka, and Reginald R Souleyrette. 1998. "Iowa Travel Model Performance, Twenty Years Later." In *Transportation, Land Use, and Air Quality: Making the Connection*, 586–95. ASCE.

APTA. 2020. "Ridership Report." *American Public Transportation Association* (blog). 2020. https://www.apta.com/research-technical-resources/transit-statistics/ridership-report/.

"Arup Settles $1.7B Australia Toll Road Revenue Forecast Suit." n.d. Accessed June 21, 2020. https://www.enr.com/articles/43707-arup-settles-17b-australia-toll-road-revenue-forecast-suit.

Ascher, William. 1979. *Forecasting: An Appraisal for Policy-Makers and Planners*.

Australia Government. 2012. "Addressing Issues in Patronage Forecasting for PPP/Toll Roads." Department of Infrastructure, Regional Development and Cities Canberra, Australia. https://infrastructure.gov.au/infrastructure/public_consultations/files/patronage_consultation_paper.pdf.

Bain, Robert. 2009. "Error and Optimism Bias in Toll Road Traffic Forecasts." *Transportation* 36 (5): 469–82. https://doi.org/10.1007/s11116-009-9199-7.

Bain, Robert 2011. "On the Reasonableness of Traffic Forecasts." *TEC Magazine*, May 2011.

Bain, Robert 2013. "Toll Roads: Big Trouble Down Under." *Infrastructure Journal*, January 17, 2013.

Bain, Robert, and Lidia Polakovic. 2005. "Traffic Forecasting Risk Study Update 2005: Through Ramp-up and Beyond." *Standard & Poor's, London*. http://toolkit.pppinindia.com/pdf/standard-poors.pdf.

Barnes, Michelle L., and Anthony (Tony) W. Hughes. 2002. "A Quantile Regression Analysis of the Cross Section of Stock Market Returns." *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.458522.

Bastian, Anne, Maria Börjesson, and Jonas Eliasson. 2016. "Explaining 'Peak Car' with Economic Variables." *Transportation Research Part A: Policy and Practice* 88 (June): 236–50. https://doi.org/10.1016/j.tra.2016.04.005.

Berrebi, Simon J, Taylor Gibbs, Sanskruti Joshi, and Kari E Watkins. In Review. "On Ridership and Frequency," 26. https://arxiv.org/abs/2002.02493.

Berrebi, Simon J., and Kari E. Watkins. 2020. "Who's Ditching the Bus?" *Transportation Research Part A: Policy and Practice* 136 (June): 21–34. https://doi.org/10.1016/j.tra.2020.02.016.

Boisjoly, Geneviève, Emily Grisé, Meadhbh Maguire, Marie-Pier Veillette, Robbin Deboosere, Emma Berrebi, and Ahmed El-Geneidy. 2018. "Invest in the Ride: A 14 year Longitudinal Analysis of the Determinants of Public Transport Ridership in 25 North American Cities." *Transportation Research Part A: Policy and Practice* 116 (October): 434–45. https://doi.org/10.1016/j.tra.2018.07.005.

Boyle, Daniel, Transportation Research Board, and Transit Cooperative Research Program Synthesis Program. 2006. *TCRP Synthesis 66: Fixed-Route Transit Ridership Forecasting and Service Planning Methods*.

Brakewood, Candace, Gregory S. Macfarlane, and Kari Watkins. 2015. "The Impact of Real-Time Information on Bus Ridership in New York City." *Transportation Research Part C: Emerging Technologies* 53 (April): 59–75. https://doi.org/10.1016/j.trc.2015.01.021.

Briggs, Andrew H., Milton C. Weinstein, Elisabeth A. L. Fenwick, Jonathan Karnon, Mark J. Sculpher, and A. David Paltiel. 2012. "Model Parameter Estimation and Uncertainty: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6." *Value in Health* 15 (6): 835–42. https://doi.org/10.1016/j.jval.2012.04.014.

Brooks, Michael. 2019. *Planning Theory for Practitioners*. Routledge.

Brown, Jeffrey, and Dristi Neog. 2012. "Central Business Districts and Transit Ridership: A Reexamination of the Relationship in the United States." *Journal of Public Transportation* 15 (4). https://doi.org/10.5038/2375-0901.15.4.1.

BTS. n.d. "Locations of Docked Bikeshare Stations by System and Year." Bureau of Transportation Statistics. Accessed February 10, 2021. https://data.bts.gov/Bicycles-and-Pedestrians/Locations-of-Docked-Bikeshare-Stations-by-System-a/7m5x-ubud.

Buck, Darren, Ralph Buehler, Patricia Happ, Bradley Rawls, Payton Chung, and Natalie Borecki. 2013. "Are Bikeshare Users Different from Regular Cyclists?: A First Look at Short-Term Users, Annual Members, and Area Cyclists in the Washington, D.C., Region." *Transportation Research Record* 2387 (1): 112–19. https://doi.org/10.3141/2387-13.

Buck, Karl, and Mike Sillence. 2014a. "A Review of the Accuracy of Wisconsin's Traffic Forecasting Tools." In Transportation Research Board 93rd Annual Meeting. https://trid.trb.org/view/2014/C/1287942.

Button, Kenneth J., Soogwan Doh, Matthew H. Hardy, Junyang Yuan, and Xin Zhou. 2010. "The Accuracy Of Transit System Ridership Forecasts And Capital Cost Estimates." *International Journal of Transport Economics / Rivista Internazionale Di Economia Dei Trasporti* 37 (2): 155–68. https://www.jstor.org/stable/42747901.

Byala, Lora B, Katie Filardo, Oren Hirsch, Michael J Walk, James P Cardenas, and Jinuk Hwang. 2019. "Comprehensive Bus Network Redesigns."

Cade, Brian S., and Barry R. Noon. 2003. "A Gentle Introduction to Quantile Regression for Ecologists." *Frontiers in Ecology and the Environment* 1 (8): 412–20. https://doi.org/10.1890/1540-9295(2003)001[0412:AGITQR]2.0.CO;2.

Campbell, Kayleigh B., and Candace Brakewood. 2017. "Sharing Riders: How Bikesharing Impacts Bus Ridership in New York City." *Transportation Research Part A: Policy and Practice* 100 (June): 264–82. https://doi.org/10.1016/j.tra.2017.04.017.

Chakrabarti, Sandip, and Genevieve Giuliano. 2015. "Does Service Reliability Determine Transit Patronage? Insights from the Los Angeles Metro Bus System." *Transport Policy* 42: 12–20.

Chang, Justin S., Sung-bong Chung, Kyu-hwa Jung, and Ki-min Kim. 2010. "Patronage Ramp-Up Analysis Model Using a Heuristic F-Test." *Transportation Research Record* 2175 (1): 84–91. https://doi.org/10.3141/2175-10.

Chen, Cynthia, Don Varley, and Jason Chen. 2011. "What Affects Transit Ridership? A Dynamic Analysis Involving Multiple Factors, Lags and Asymmetric Behaviour." *Urban Studies* 48 (9): 1893–1908. https://www.jstor.org/stable/43081823.

Clewlow, Regina R. 2019. "The Micro-Mobility Revolution: The Introduction and Adoption of Electric Scooters in the United States." In . Washington DC, United States. https://trid.trb.org/view/1572549.

Collaboration for Nondestructive Testing. n.d. "Uncertainty Terminology." Accessed July 2,2020.https://www.ndeed.org/GeneralResources/ErrorAnalysis/UncertaintyTerms.htm.

Davidson, C. 2014. "Future Demand–Insights into the Scenario Planning Methodology." *New Zealand Ministry of Transport, Wellington, New Zealand, November. Available (as at 16/02/2016) From< Www. Transport. Govt. Nz/Futuredemand.*

Davis, Jeff. 2019. “U.S. VMT Per Capita By State, 1981-2017.” *Eno Center for Transportation* (blog). June 7, 2019. https://www.enotrans.org/eno-resources/u-s-vmt-per-capita-by-state-1981-2017/.

Delgadillo, Natalie. 2020. “Metro Saw Its First Ridership Increase In Nearly A Decade During 2019.” *WAMU*, January 16, 2020. https://wamu.org/story/20/01/16/metro-saw-its-first-ridership-increase-in-nearly-a-decade-during-2019/.

Dill, Jennifer. 2013. “Predicting Transit Ridership At The Stop Level: The Role Of Service And Urban Form,” In 92nd annual meeting of the Transportation Research Board, Washington, DC.

Dill, Jennifer, Marc Schlossberg, Liang Ma, and Cody Meyer. 2013. “Predicting Transit Ridership at the Stop Level: The Role of Service and Urban Form.” In *92nd Annual Meeting of the Transportation Research Board, Washington, DC*.

Driscoll, Richard A., Kurt R. Lehmann, Steven Polzin, and Jodi Godfrey. 2018. “The Effect of Demographic Changes on Transit Ridership Trends.” *Transportation Research Record* 2672 (8): 870–78. https://doi.org/10.1177/0361198118777605.

Duggan, Paul, Lori Aratani, and Robert McCartney. 2016. “Metro Will Shut down Sections of Lines for Year-Long Subway Repair Work.” *Washington Post*, May 6, 2016, sec.Transportation. https://www.washingtonpost.com/local/trafficandcommuting/metro-armageddon-is-coming-a-year-of-shuttered-sections-massive-repairs/2016/05/06/9a7160d0-12e6-11e6-93ae-50921721165d_story.html.

Ederer, David, Simon Berrebi, Chandler Diffee, Taylor Gibbs, and Kari E Watkins. 2019a. “Comparing Transit Agency Peer Groups Using Cluster Analysis.” *Transportation Research Record*, 0361198119852074.

Ederer, David, Simon Berrebi, Chandler Diffee, Taylor Gibbs, and Kari E Watkins. 2019b. “Comparing Transit Agency Peer Groups Using Cluster Analysis.” *Transportation Research Record* 2673 (11): 505–16.

Eliasson, Jonas, and Mogens Fosgerau. 2013. “Cost Overruns and Demand Shortfalls – Deception or Selection?” *Transportation Research Part B: Methodological* 57: 105–13. https://doi.org/10.1016/j.trb.2013.09.005.

Erhardt, Greg, David Schmitt, Jawad Hoque, Ankita Chaudhary, Sujith Rapolu, Kyeongsu Kim, Steve Weller, et al. 2020. “Traffic Forecasting Accuracy Assessment Research.” National Cooperative Highway Research Program NCHRP Report 934. Transportation Research Board. https://doi.org/10.17226/25637.

Erhardt, Gregory D. 2017. “Conceptual Models of the Effect of Information and Communications Technology on Long-Distance Travel Demand.” *Transportation Research Record: Journal of the Transportation Research Board* 2658: 26–34. https://doi.org/10.3141/2658-04.

Erhardt, Gregory D, Jawad Hoque, Mei Chen, Reginald Souleyrette, David Schmitt, Ankita Chaudhary, Sujith Rapolu, et al. 2019. "Traffic Forecasting Accuracy Assessment Research." *NCHRP Research Report*, no. 934.

Erhardt, Gregory D., Jawad Mahmud Hoque, Vedant Goyal, Simon Berrebi, Candace Brakewood, and Kari Edison Watkins. Submitted. "Why Has Transit Ridership Declined in the US." *Transportation Research Part A: Policy and Practice*.

Erhardt, Gregory D., Richard Alexander Mucci, Drew Cooper, Bhargava Sana, Mei Chen, and Joe Castiglione. 2021. "Do Transportation Network Companies Increase or Decrease Transit Ridership? Empirical Evidence from San Francisco." *Transportation*, February. https://doi.org/10.1007/s11116-021-10178-4.

Evans IV, John E. 2004. "Transit Scheduling and Frequency-Traveler Response to Transportation System Changes."

Federal Transit Administration. 2020. "Predicted Versus Actual Impacts of Capital Investment Grants Projects - 2020." 2020. Federal Transit Administration.

Florida, Richard. 2017. *The New Urban Crisis: How Our Cities Are Increasing Inequality, Deepening Segregation, and Failing the Middle Class-and What We Can Do About It*. Basic Books.

Flyvbjerg, B., M. K. S. Holm, and S. L. Buhl. 2005. "How (In)Accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation." *Journal of the American Planning Association* 71 (2). https://trid.trb.org/view.aspx?id=755586.

Flyvbjerg, Bent. 2005. "Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations Regarding Ramp up and Sampling." *Transportation Research Part A: Policy and Practice* 39 (6): 522–30. https://doi.org/10.1016/j.tra.2005.02.003.

Flyvbjerg, Bent. 2007. "Policy and Planning for Large-Infrastructure Projects: Problems, Causes, Cures." *Environment and Planning B: Planning and Design* 34 (4): 578–97. https://doi.org/10.1068/b32111.

Flyvbjerg, Bent. 2013. "Quality Control and Due Diligence in Project Management: Getting Decisions Right by Taking the Outside View." *International Journal of Project Management* 31 (5): 760–74.

Flyvbjerg, Bent, M. K. S. Holm, and S. L. Buhl. 2005. "How (In)Accurate Are Demand Forecasts in Public Works Projects? The Case of Transportation." *Journal of the American Planning Association* 71 (2). https://trid.trb.org/view.aspx?id=755586.

Flyvbjerg, Bent, Skamris Holm, M. K, and S. L. Buhl. 2006. "Inaccuracy in Traffic Forecasts." *Transport Reviews* 26 (1). https://trid.trb.org/view/2006/C/781962.

Flyvbjerg, Bent, Mette K Skamris Holm, and Søren L Buhl. 2005. "How (in) Accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation." *Journal of the American Planning Association* 71 (2): 131–46.

Frey, William H. 2018. “US Population Disperses to Suburbs, Exurbs, Rural Areas, and ‘Middle of the Country’ Metros.” *Brookings Institution*.

Fuller, Daniel, Lise Gauvin, Yan Kestens, Patrick Morency, and Louis Drouin. 2013. “The Potential Modal Shift and Health Benefits of Implementing a Public Bicycle Share Program in Montreal, Canada.” *International Journal of Behavioral Nutrition and Physical Activity* 10 (1): 66. https://doi.org/10.1186/1479-5868-10-66.

Giaimo, Greg, and Mark Byram. 2013. “Improving Project Level Traffic Forecasts by Attacking the Problem from All Sides.” Presented at the The 14th Transportation Planning Applications Conference, Columbus, OH.

Gomez-Ibanez, Jose A. 1996. “Big-City Transit Rider Snip, Deficits, and Politics: Avoiding Reality in Boston.” *Journal of the American Planning Association* 62 (1): 30–50.

Graehler, Michael, Richard Alexander Mucci, and Gregory D. Erhardt. 2019. “Understanding the Recent Transit Ridership Decline in Major US Cities: Service Cuts or Emerging Modes?” In *Transportation Research Board Annual Meeting*. Washington, D.C.

Grimsrud, Michael, and Ahmed El-Geneidy. 2013. “Driving Transit Retention to Renaissance: Trends in Montreal Commute Public Transport Mode Share and Factors by Age Group and Birth Cohort.” *Public Transport* 5 (3): 219–41. https://doi.org/10.1007/s12469-013-0075-7.

Grimsrud, Michael, and Ahmed El-Geneidy. 2014. “Transit to Eternal Youth: Lifecycle and Generational Trends in Greater Montreal Public Transport Mode Share.” *Transportation* 41 (1): 1–19. https://doi.org/10.1007/s11116-013-9454-9.

Hall, Jonathan D., Craig Palsson, and Joseph Price. 2018. “Is Uber a Substitute or Complement for Public Transit?” *Journal of Urban Economics* 108 (November): 36–50. https://doi.org/10.1016/j.jue.2018.09.003.

Hartgen, David T. 2013. “Hubris or Humility? Accuracy Issues for the next 50 Years of Travel Demand Modeling.” *Transportation* 40 (6): 1133–57. https://doi.org/10.1007/s11116-013-9497-y.

Hendrickson, Chris. 1986. “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): 33–37. https://doi.org/10.1016/0191-2607(86)90013-0.

Hoque, Jawad Mahmud, Gregory D Erhardt, David Schmitt, Mei Chen, Ankita Chaudhary, Martin Wachs, and Reginald R Souleyrette. 2021a. “The Changing Accuracy of Traffic Forecasts.” *Transportation*, 1–22.

Hoque, Jawad Mahmud, Gregory D. Erhardt, David Schmitt, Mei Chen, Ankita Chaudhary, Martin Wachs, and Reginald R. Souleyrette. 2021b. “The Changing Accuracy of Traffic Forecasts.” *Transportation*, February. https://doi.org/10.1007/s11116-021-10182-8.

Hoque, Jawad Mahmud, Gregory D Erhardt, David Schmitt, Mei Chen, and Martin Wachs. 2021. "Estimating the Uncertainty of Traffic Forecasts from Their Historical Accuracy." *Transportation Research Part A: Policy and Practice* 147: 339–49.

Horowitz, Alan, Tom Creasey, Ram Pendyala, and Mei Chen. 2014. *Analytical Travel Forecasting Approaches for Project-Level Planning and Design*. Project 08-83. Transportation Research Board.

Hugosson, Muriel Beser. 2005. "Quantifying Uncertainties in a National Forecasting Model." *Transportation Research Part A: Policy and Practice* 39 (6): 531–47. https://doi.org/10.1016/j.tra.2005.02.010.

Ismart, Dane. 1990. *Calibration and Adjustment of System Planning Models*. Federal Highway Administration.

ISO 5725-1. 1994. "Accuracy (Trueness and Precision) of Measurement Methods and Results — Part 1: General Principles and Definitions." https://www.iso.org/obp/ui/#iso:std:iso:5725:-1:ed-1:v1:en.

ISO/IEC Guide 98-3. 2008. "ISO/IEC Guide 98-3:2008(En), Uncertainty of Measurement — Part 3: Guide to the Expression of Uncertainty in Measurement (GUM:1995)." https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:98:-3:ed-1:v2:en.

Jong, Gerard de, Andrew Daly, Marits Pieters, Stephen Miller, Ronald Plasmeijer, and Frank Hofman. 2007. "Uncertainty in Traffic Forecasts: Literature Review and New Results for The Netherlands." *Transportation* 34 (4): 375–95. https://doi.org/10.1007/s11116-006-9110-8.

Kahneman, Daniel, and Amos Tversky. 1977. "Intuitive Prediction: Biases and Corrective Procedures." DTIC Document. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA047747.

Kain, John F. 1990. "Deception in Dallas: Strategic Misrepresentation in Rail Transit Promotion and Evaluation." *Journal of the American Planning Association* 56 (2): 184–96. http://www.honolulutraffic.com/kaindall.pdf.

Kittelson & Associates, Parsons Brinckerhoff, KFH Group, Inc., Texas A&M Transportation Institute, and Arup. 2013. "Transit Capacity and Quality of Service Manual." Transit Cooperative Research Program TCRP Report 165. Washington, D.C: Transportation Research Board. http://www.trb.org/TCRP/Blurbs/169437.aspx.

Knaap, Gerrit-Jan, Daniel Engelberg, Uri Avin, Sevgi Erdogan, Fred Ducca, Timothy F. Welch, Nicholas Finio, Rolf Moeckel, and Harutyun Shahumyan. 2020. "Modeling Sustainability Scenarios in the Baltimore–Washington (DC) Region." *Journal of the American Planning Association* 86 (2): 250–63. https://doi.org/10.1080/01944363.2019.1680311.

Koenker, Roger, Stephen Portnoy, Pin Tian Ng, Achim Zeileis, Philip Grosjean, Cleve Moler, and Brian D. Ripley. 2018. *Quantreg: Quantile Regression* (version 5.36). https://CRAN.R-project.org/package=quantreg.

Kohn, H. 2000. "Factors Affecting Urban Transit Ridership: Statistics Canada." *Transportation Division*.

Kriger, David, Suzette Shiu, and Sasha Naylor. 2006. "Estimating Toll Road Demand and Revenue." Synthesis of Highway Practice NCHRP Synthesis 364. Transportation Research Board. https://trid.trb.org/view/2006/M/805554.

Kyte, Michael, James Stoner, and Jonathan Cryer. 1988. "A Time-Series Analysis of Public Transit Ridership in Portland, Oregon, 1971–1982." *Transportation Research Part A: General* 22 (5): 345–59.

Lê Cook, Benjamin, and Willard G. Manning. 2013. "Thinking beyond the Mean: A Practical Guide for Using Quantile Regression Methods for Health Services Research." *Shanghai Archives of Psychiatry* 25 (1): 55–59. https://doi.org/10.3969/j.issn.1002-0829.2013.01.011.

Lee, Bumsoo, and Yongsung Lee. 2013. "Complementary Pricing and Land Use Policies: Does It Lead to Higher Transit Use?" *Journal of the American Planning Association* 79 (4): 314–28. https://doi.org/10.1080/01944363.2014.915629.

Lemp, Jason D., and Kara M. Kockelman. 2009. "Understanding and Accommodating Risk and Uncertainty in Toll Road Projects: A Review of the Literature." *Transportation Research Record* 2132 (1): 106–12. https://doi.org/10.3141/2132-12.

Lewis-Workman, Steven, Tom Buffkin, Sarah Clements, and Joe Swartz. 2003. "The Predicted and Actual Impacts of New Starts Projects." *US Department of Transportation Federal Transit Administration, Washington, DC*.

Lewis-Workman, Steven, Bryon White, Stephanie McVey, and Frank Spielberg. 2007. "The Predicted and Actual Impacts of New Starts Projects." *Washington DC: US Department of Transportation Federal Transit Administration*.

Li, Zheng, and David A. Hensher. 2010. "Toll Roads in Australia: An Overview of Characteristics and Accuracy of Demand Forecasts." *Transport Reviews* 30 (5): 541–69. https://doi.org/10.1080/01441640903211173.

Liu, Luyu, Harvey J. Miller, and Jonathan Scheff. 2020. "The Impacts of COVID-19 Pandemic on Public Transit Demand in the United States." *PLOS ONE* 15 (11): e0242476. https://doi.org/10.1371/journal.pone.0242476.

Liu, Zhi. 1993. "DETERMINANTS OF PUBLIC TRANSIT RIDERSHIP ANALYSIS OF POST WORLD WAR II TRENDS AND EVALUATION OF ALTERNATIVE NETWORKS."

Loa, Patrick, Sanjana Hossain, Yicong Liu, and Khandker Nurul Habib. 2022. "How Has the COVID-19 Pandemic Affected the Use of Ride-Sourcing Services? An Empirical Evidence-Based Investigation for the Greater Toronto Area."

Transportation Research Part A: Policy and Practice 155 (January): 46–62. https://doi.org/10.1016/j.tra.2021.11.013.

Lyons, Glenn, and Cody Davidson. 2016. "Guidance for Transport Planning and Policymaking in the Face of an Uncertain Future." *Transportation Research Part A: Policy and Practice* 88 (June): 104–16. https://doi.org/10.1016/j.tra.2016.03.012.

Lyons, Glenn, and Greg Marsden. 2019. "Opening out and Closing down: The Treatment of Uncertainty in Transport Planning's Forecasting Paradigm." *Transportation*, November. https://doi.org/10.1007/s11116-019-10067-x.

Ma, Ting, and Gerrit-Jan Knaap. 2019. "Estimating the Impacts of Capital Bikeshare on Metrorail Ridership in the Washington Metropolitan Area." *Transportation Research Record* 2673 (7): 371–79. https://doi.org/10.1177/0361198119849407.

Mahmoud, Mohamed Salah, and Mark Pickup. 2019. "Estimating Transit Fare Elasticity Using Panel Data Models–Metro Vancouver Case Study."

Maley, Donald W., and Rachel Weinberger. 2009. "Rising Gas Price and Transit Ridership: Case Study of Philadelphia, Pennsylvania." *Transportation Research Record* 2139 (1): 183–88. https://doi.org/10.3141/2139-21.

Manville, Michael, Brian D. Taylor, and Evelyn Blumenberg. 2018. "Falling Transit Ridership: California and Southern California." Prepared for the Southern California Association of Governments. UCLA Institute for Transportation Studies. https://www.scag.ca.gov/Documents/ITS_SCAG_Transit_Ridership.pdf.

Manzo, Stefano, Otto Anker Nielsen, and Carlo Giacomo Prato. 2015a. "How Uncertainty in Input and Parameters Influences Transport Model Output: A Four-Stage Model Case-Study." *Transport Policy* 38 (February): 64–72. https://doi.org/10.1016/j.tranpol.2014.12.004.

Manzo, Stefano, Otto Anker Nielsen, and Carlo Giacomo Prato. 2015b. "How Uncertainty in Socio-Economic Variables Affects Large-Scale Transport Model Forecasts." *European Journal of Transport and Infrastructure Research* 15 (3). https://doi.org/10.18757/ejtir.2015.15.3.3080.

Marlin Engineering. 2015. "Traffic Forecasting Sensitivity Analysis." TWO#13.

Martin, Elliot, and Susan Shaheen. 2011. "The Impact of Carsharing on Public Transit and Non-Motorized Travel: An Exploration of North American Carsharing Survey Data." *Energies* 4 (11): 2094–2114. https://doi.org/10.3390/en4112094.

Mason, Simon J, and David B Stephenson. 2008. "How Do We Know Whether Seasonal Climate Forecasts Are Any Good?" In *Seasonal Climate: Forecasting and Managing Risk*, 259–89. Springer.

Meyer, Michael D, and Eric J Miller. 2001. *Urban Transportation Planning: A Decision-Oriented Approach*. 2nd ed.

Miller, John S., Salwa Anam, Jasmine W. Amanin, and Raleigh A. Matteo. 2016. "A Retrospective Evaluation of Traffic Forecasting Techniques." FHWA/VTRC 17-R1. Virginia Transportation Research Council. http://ntl.bts.gov/lib/37000/37800/37804/10-r24.pdf.

Mintzberg, Henry. 1981. "What Is Planning Anyway?" *Strategic Management Journal* 2 (3): 319–24. http://www.jstor.org/stable/2486227.

Murphy, Allan H. 1993. "What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting." *Weather and Forecasting* 8 (2): 281–93.

NACTO. 2018. "Shared Micromobility in the U.S.: 2018." Www.Nacto.Org. 2018. https://nacto.org/wp-content/uploads/2019/04/NACTO_Shared-Micromobility-in-2018_Web.pdf.

National Academies of Sciences, Engineering, and Medicine. 2018. "Understanding Changes in Demographics, Preferences, and Markets for Public Transportation." *World Transit Research*, January. https://www.worldtransitresearch.info/research/7101.

Naylor, Thomas H. 1983. "Strategic Planning and Forecasting." *Journal of Forecasting* 2 (2): 109–18. https://doi.org/10.1002/for.3980020203.

Nicolaisen, Morten Skou. 2012. "Forecasts: Fact or Fiction?: Uncertainty and Inaccuracy in Transport Project Evaluation." PhD Thesis, Videnbasen for Aalborg UniversitetVBN, Aalborg UniversitetAalborg University, Det Teknisk-Naturvidenskabelige FakultetThe Faculty of Engineering and Science. http://vbn.aau.dk/ws/files/81035038/Nicolaisen_2012_Forecasts_Fact_or_Fiction_PhD_Thesis.pdf.

Nicolaisen, Morten Skou, and Patrick Arthur Driscoll. 2014. "Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review." *Transport Reviews* 34 (4): 540–57. https://doi.org/10.1080/01441647.2014.926428.

Nowak, William P., and Ian Savage. 2013. "The Cross Elasticity between Gasoline Prices and Transit Use: Evidence from Chicago." *Transport Policy* 29 (September): 38–45. https://doi.org/10.1016/j.tranpol.2013.03.002.

Odeck, James, and Morten Welde. 2017. "The Accuracy of Toll Road Traffic Forecasts: An Econometric Evaluation." *Transportation Research Part A: Policy and Practice* 101 (July): 73–85. https://doi.org/10.1016/j.tra.2017.05.001.

Parthasarathi, Pavithra, and David Levinson. 2010. "Post-Construction Evaluation of Traffic Forecast Accuracy." *Transport Policy* 17 (6): 428–43. https://doi.org/10.1016/j.tranpol.2010.04.010.

Paul, Jesse. 2021. "Driver Shortage Increases Prices for Ride-Hailing Services." *Associated Press*, August 7, 2021, sec. Health. https://apnews.com/article/health-coronavirus-pandemic-39526365faa33c5bb71272d2e11397fe.

Pedersen, N.J., and D.R. Samdahl. 1982. "Highway Traffic Data for Urbanized Area Project Planning and Design." National Cooperative Highway Research Program NCHRP 255. Washington, D.C.: Transportation Research Board.

Pereira, Francisco C., Constantinos Antoniou, Joan Aguilar Fargas, and Moshe Ben-Akiva. 2014. "A Metamodel for Estimating Error Bounds in Real-Time Traffic Prediction Systems." *IEEE Transactions on Intelligent Transportation Systems* 15 (3): 1310–22. https://doi.org/10.1109/TITS.2014.2300103.

Perk, Victoria A, Nilgun Kamp, Melissa Salzler, and others. 2004. "Benchmark Rankings for Transit Systems in the United States." University of South Florida. Center for Urban Transportation Research.

Pickrell, D. H. 1990. "Urban Rail Transit Projects: Forecast Versus Actual Ridership and Costs. Final Report." https://trid.trb.org/view.aspx?id=299240.

"President Biden's Bipartisan Infrastructure Law." n.d. The White House. Accessed March 14, 2022. https://www.whitehouse.gov/bipartisan-infrastructure-law/.

Schaller, Bruce. 2018. "The New Automobility: Lyft, Uber and the Future of American Cities," July. https://trid.trb.org/view/1527868.

Schmitt, Angie. 2017. "Transit Ridership Falling Everywhere — But Not in Cities With Redesigned Bus Networks." *Streetsblog USA* (blog). February 24, 2017. https://usa.streetsblog.org/2017/02/24/transit-ridership-falling-everywhere-but-not-in-cities-with-redesigned-bus-networks/.

Schmitt, David. 2016. "A Transit Forecasting Accuracy Database: Beginning to Enjoy the 'Outside View.'" In *Transportation Research Board 95th Annual Meeting*. Washington, D.C.

Schrank, David, Bill Eisele, and Tim Lomax. 2019. "2019 Urban Mobility Report." Texas A&M Transportation Institute.

Shaheen, Susan A, Elliot W Martin, Adam P Cohen, Nelson D Chan, and Mike Pogodzinski. 2014. "Public Bikesharing in North America During a Period of Rapid Expansion: Understanding Business Models, Industry Trends & User Impacts, MTI Report 12-29," 236.

Shinn, Jill Elizabeth, and Carole Turley Voulgaris. 2019. "Ridership Ramp-Up? Initial Ridership Variation on New Rail Transit Projects:" *Transportation Research Record*, May. https://doi.org/10.1177/0361198119844462.

Stanley, Robert. 1998. "Continuing Examination of Successful Transit Ridership Initiatives." *TCRP Reasearch Digest*, 29, , 68.

Steiner, George A. 2010. *Strategic Planning*. Simon and Schuster.

Tang, Lei, and Piyushimita Thakuriah. 2012. "Ridership Effects of Real-Time Bus Information System: A Case Study in the City of Chicago." *Transportation Research Part C: Emerging Technologies* 22: 146–61. https://doi.org/10.1016/j.trc.2012.01.001.

Taylor, Brian D., Douglas Miller, Hiroyuki Iseki, and Camille Fink. 2009. “Nature and/or Nurture? Analyzing the Determinants of Transit Ridership across US Urbanized Areas.” *Transportation Research Part A: Policy and Practice* 43 (1): 60–77. https://doi.org/10.1016/j.tra.2008.06.007.

Taylor, Brian D., and Jacob L. Wasserman. 2021. “Sources of and Gaps in Data for Understanding Public Transit Ridership.” UCLA Research Report UC-ITS-2020-33. https://escholarship.org/uc/item/22n525dc.

“The National Transit Database (NTD) | FTA.” n.d. Accessed November 24, 2020. https://www.transit.dot.gov/ntd.

“The Transport Politic - Transit Explorer 2021.” n.d. Accessed July 19, 2021. https://www.thetransportpolitic.com/transitexplorer/.

Thistle, Ian, and Alissa Zimmer. 2019. “Location, Location, Location: A Neighborhood-Level Analysis of Changes in MBTA Ridership.” MBTA – Office of Performance Management and Innovation, Boston. https://massdot.box.com/v/busridershipreport.

Transportation Research Board, and Engineering National Academies of Sciences and Medicine. 2010. *Legal Handbook for the New Starts Process*. Washington, DC: The National Academies Press. https://doi.org/10.17226/22970.

US Census Bureau. 2012. “How Do We Measure Urban Areas?” The United States Census Bureau. 2012. https://www.census.gov/newsroom/blogs/random-samplings/2012/04/how-do-we-measure-urban-areas.html.

U.S. Department of Transportation, Federal Highway Administration. n.d. “Public Information about the ‘FAST Act.’” Accessed October 6, 2016. https://www.fhwa.dot.gov/fastact/funding.cfm.

US Securities and Exchange Commission. 2019a. “Form S-1 Registration Satement: Lyft Inc.” https://www.sec.gov/Archives/edgar/data/1759509/000119312519059849/d633517ds1.htm.

US Securities and Exchange Commission. 2019b. “Form S-1 Registration Satement: Uber Technologies, Inc.” https://www.sec.gov/Archives/edgar/data/1543151/000119312519103850/d647752ds1.htm.

Voulgaris, Carole Turley. 2019a. “Trust in Forecasts? Correlates with Ridership Forecast Accuracy for Fixed-Guideway Transit Projects.” *Transportation*, 1–39.

Voulgaris, Carole Turley. 2019b. “Crystal Balls and Black Boxes: What Makes a Good Forecast?” *Journal of Planning Literature* 34 (3): 286–99. https://doi.org/10.1177/0885412219838495.

Wachs, M. 1985. “Ethics in Planning. New Brunswick, New Jersey: Center for Urban Policy Research.”

Wachs, Martin. 1990. “Ethics and Advocacy in Forecasting for Public Policy.” *Business and Professional Ethics Journal* 9 (1 & 2): 141–57.

Watkins, Kari, Simon Berrebi, Chandler Diffee, Becca Kiriazes, and David Ederer. 2020. “Analysis of Recent Public Transit Ridership Trends.” Transit Cooperative Research Program TCRP Report 209. Washington, D.C.: Transportation Research Board.

Watkins, Kari, Simon Berrebi, Gregory Erhardt, Jawad Hoque, Vedant Goyal, Candace Brakewood, Abubakr Ziedan, Brendon Hemily, and Josie Kressner. 2021. “Recent Decline in Public Transportation Ridership: Analysis, Causes, Responses.” Pre-Publication Draft TCRP Research Report 231. Transportation Research Board. https://doi.org/10.17226/26320.

Webber, Melvin W. 1976. “The BART Experience—What Have We Learned?” *Institute of Urban & Regional Development*, October. http://escholarship.org/uc/item/7pd9k5g0.

Welde, Morten, and James Odeck. 2011. “Do Planners Get It Right? The Accuracy of Travel Demand Forecasting in Norway.” *EJTIR* 1 (11): 80–95. https://doi.org/10.18757/ejtir.2011.11.1.2913.

Yanmaz-Tuzel, Ozlem, and Kaan Ozbay. 2010. “Impacts of Gasoline Prices on New Jersey Transit Ridership.” *Transportation Research Record* 2144 (1): 52–61. https://doi.org/10.3141/2144-07.

Zhang, Xu, and Mei Chen. 2017. “Quantifying Effects from Weather on Travel Time and Reliability.” In , 14. Washington, D.C.

Zhang, Xu, and Mei Chen. 2017. 2019. “Quantifying the Impact of Weather Events on Travel Time and Reliability.” *Journal of Advanced Transportation*. https://doi.org/10.1155/2019/8203081.

Zhao, Yong, and Kara Maria Kockelman. 2002. “The Propagation of Uncertainty through Travel Demand Models: An Exploratory Analysis.” *The Annals of Regional Science* 36 (1): 145–63. https://doi.org/10.1007/s001680200072.

Ziedan, Abubakr, Wesley Darling, Candace Brakewood, Greg Erhardt, and Kari Watkins. 2021. “The Impacts of Shared E-Scooters on Bus Ridership.” *Transportation Research Part A: Policy and Practice* 153 (November): 20–34. https://doi.org/10.1016/j.tra.2021.08.019.

VITA

Jawad Mahmud Hoque

Ph.D. Candidate

jawadmhoque@uky.edu

EDUCATION

Ph.D.	University of Kentucky	Expected May 2022
	Civil Engineering	
	Dissertation: Assessing and Quantifying the Uncertainty in Traffic and Transit Ridership Forecasts	
M.Sc.	University of Kentucky	May 2019
	Civil Engineering	
	Thesis: An Assessment of Historical Traffic Forecast Accuracy and Sources of Forecast Error	
B.Sc.	Bangladesh University of Engineering and Technology	April 2016

RESEARCH EXPERIENCE

Urban Transportation and Logistics Intern **May 2021-June 2021**

South Asian Region, Transportation Unit, The World Bank

Project: Auto-Trip Emission Characteristics by Households in the Greater Mumbai Region

Graduate Research Assistant **August 2017-April 2022**

Department of Civil Engineering, University of Kentucky

Project: Accuracy and Uncertainty in Transit Ridership Forecasts 2020-2022

Project: Multi-Modal Optimization and Simulation Track, 2020-2022

Project: Recent Decline in Public Transportation Ridership: Analysis, Causes, Responses — 2019-2020

Transit Cooperative Research Program (TCRP), Transportation Research Board

Project: Traffic Forecast Accuracy Assessment Research — 2017-2019

National Cooperative Highway Research Program (NCHRP), Transportation Research Board

PUBLICATIONS

Peer Reviewed Journal Articles

Hoque, J.M., Erhardt, G.D., Schmitt, D., Chen, M., Chaudhary, A., Wachs, M., Souleyrette, R. (2021) "The Changing Accuracy of Traffic Forecasts", *Transportation*.

Hoque, J.M., Erhardt, G.D., Schmitt, D., Chen, M., Wachs, M., (2021) "Estimating the Uncertainty of Traffic Forecasts from Their Historical Accuracy", *Transportation Research Part A: Policy and Practice*.

Manuscripts Under Review

Erhardt, G.D., Hoque, J.M., Watkins, K., Brakewood, C., Berrebi. S., (in review) "Why has public transit ridership declined in the United States?"*, Transportation Research Part A: Policy and Practice*.

Research Reports

Watkins, K., Berrebi, S., Erhardt, G.D., Hoque, J., Goyal, V., Brakewood, C., Ziedan, A., Hemily, B., Kressner, J. (2022), *Recent Decline in Public Transportation Ridership: Analysis, Causes, Responses,* Transit Cooperative Research Program, Transportation Research Board, Washington, D.C.

Erhardt, G.D., Schmitt, D., Hoque, J., Chaudhary, A., Rapolu, S., Kim, K., Weller, S., Sall, E., Chen, M., Souleyrette, R., Wachs, M. (2020), *Traffic Forecasting Accuracy Assessment Research*, National Cooperative Highway Research Program Report 934, Transportation Research Board, Washington, D.C.

PRESENTATIONS AND POSTERS

Conference Participation

Hoque, J.M, Schmitt, D., Erhardt, G.D., "Quantifying Uncertainty in Traffic Forecasts: A Retrospective Approach", Transportation Research Board National Transportation Planning Applications Conference, June 2021.

Schmitt, D., Erhardt, G.D., Hoque, J.M, "Interim Findings from NCHRP Project 8-110, Traffic Forecasting Accuracy", speaker and interactive panelist at Workshop 1061 on "Progress in Improving Travel Forecasting Accuracy", 98th Transportation Research Board Annual Meeting, Washington, D.C., January 2019.

Webinars

Erhardt, G.D., Schmitt, D., Hoque, J., Chen, M. "How Accurate are Traffic Forecasts?", Transportation Research Board Webinar, October 2020.

Hoque, J., "Prospective and Retrospective Views of Uncertainty in Travel Forecasting", Zephyr Virtual Sessions (online), June 2020.

Poster Sessions

Hoque, J.M., Erhardt, G. "Quantifying the Inherent Uncertainty in Traffic Forecasts Using Quantile Regression Methods," poster presented by Jawad Hoque at the Commonwealth Computational Summit, Lexington, Kentucky, October 2019.

Plenary Sessions

Schmitt, D., Hoque, J., Sall, E., Erhardt, G.D., "Forecasting Accuracy: Implementing New Practices", presented by David Schmitt, Jawad Hoque and Elizabeth Sall at the closing plenary session of the Transportation Research Board National Transportation Planning Applications Conference, Portland, Oregon, June 2019.

AWARDS AND SCHOLARSHIPS

USEC Inc. Graduate Fellowship	2021
College of Engineering, University of Kentucky	
Conference Travel Grant	2021
Graduate Student Council, University of Kentucky	
William (Bill) Seymour Transportation Engineering Scholarship	2020
Kentucky Section of Institute of Transportation Engineers (KYSITE)	
William H. Temple Scholarship	2019
Champion of the Southern District Institute of Transportation Engineers (SDITE) Traffic Bowl	
William H. Temple Scholarship	2018
Runners up of the Southern District Institute of Transportation Engineers (SDITE) Traffic Bowl	

LEADERSHIP

Chair May 2021-March 2022

Technical Session Arrangement Committee, Southern District Institute of Transportation Engineers Annual Meeting 2022

President July 2019-June 2020

Institute of Transportation Engineers, University of Kentucky Student Chapter

Vice-President August 2018- June 2019

Institute of Transportation Engineers, University of Kentucky Student Chapter