2019

An Assessment of Historical Traffic Forecast Accuracy and Sources of Forecast Error

Jawad Mahmud Hoque
University of Kentucky, jawadmhq@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0002-8803-3899
Digital Object Identifier: https://doi.org/10.13023/etd.2019.129

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Jawad Mahmud Hoque, Student

Dr. Gregory D. Erhardt, Major Professor

Dr. Timothy Taylor, Director of Graduate Studies

AN ASSESSMENT OF HISTORICAL TRAFFIC FORECAST ACCURACY AND SOURCES OF FORECAST ERROR

THESIS

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science in Civil Engineering in the
College of Engineering
at the University of Kentucky

By

Jawad Mahmud Hoque

Lexington, Kentucky

Director: Dr. Gregory D. Erhardt, Professor of Civil Engineering

Lexington, Kentucky

2019

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

ABSTRACT OF THESIS

AN ASSESSMENT OF HISTORICAL TRAFFIC FORECAST ACCURACY AND SOURCES OF FORECAST ERROR

Transportation infrastructure improvement projects are typically huge and have significant economic and environmental effects. Forecasts of demand of the facility in the form of traffic level help size the project as well as choose between several alternatives. Inaccuracy in these forecasts can thus have a great impact on the efficiency of the operational design and the benefits accrued from the project against the cost. Despite this understanding, evaluation of traffic forecast inaccuracy has been too few, especially for un-tolled roads in the United States. This study, part of a National Cooperative Highway Research Program (NCHRP) funded project, bridges this gap in knowledge by analyzing the historical inaccuracy of the traffic forecasts based on a database created as part of the project. The results show a general over-prediction of traffic with actual traffic deviating from forecast by about 17.29% on an average. The study also compares the relative accuracy of forecasts on several categorical variables. Besides enumerating the error in forecasts, this exploration presents the potential factors influencing accuracy. The results from this analysis can help create an uncertainty window around the forecast based on the explanatory variables, which can be an alternate risk analysis technique to sensitivity testing.

KEYWORDS: Traffic forecast accuracy, optimism bias in traffic forecast, distribution of forecast error, sources of forecast error.

JAWAD MAHMUD HOQUE

(Name of Student)

03/25/2019

Date

AN ASSESSMENT OF HISTORICAL TRAFFIC FORECAST ACCURACY AND SOURCES OF FORECAST ERROR

By

Jawad Mahmud Hoque

DR. GREGORY ERHARDT

Director of Thesis

DR. TIMOTHY TAYLOR

Director of Graduate Studies

03/19/2019

Date

To the three most important women in my life,
My mother,
My sister, Shemonty
And
My wife, Riana

ACKNOWLEDGMENTS

I would like to begin by thanking the Almighty, the most just and kind. Days and nights of staring blankly at the jumble of information all while the responsibilities of an "adult" life in a new country catching up to me have made me embrace my faith more and more. I am eternally grateful to Him for giving me the strength to pull through.

It has been a very interesting couple of months of riding the ebbs and flows of graduate life. I was apprehensive in the beginning, but as time passed, I started enjoying it. For this, I must thank my advisor and committee chair, Dr. Greg Erhardt. He has been my advisor in the truest meaning of the word, in all things professional, academic and even personal. I am forever grateful for the opportunity to work on this exciting research project and learning.

Next, I would like to thank the other members of my committee and collaborators in the research, Dr. Mei Chen and Dr. Reginald Souleyrette. They have provided invaluable guidance during the research work: Dr. Chen lending her immeasurable knowledge with statistical analysis and Dr. Souleyrette with his vast transportation modeling experience.

I would also like to acknowledge Dave Schmidt who provided sage advice on the utility of the research and challenged me with questions that made this research better. Thanks to Ankita Chowdhury who initialized the database. Thanks go to the state DOTs that provided the data, and to NCHRP who funded the project and the staff for provided valuable inputs.

Thanks goes to my father, for encouraging me to push my limits. Thanks to my mother, for being my strength and for getting overjoyed at the smallest achievement. Thank you, Shemonty, my sister, who is too young to understand any of this but who always kept me grounded. And thank you, Raad, my brother, for letting me rest easy while I leave my home in his hands.

But most of all, thank you, Riana, my loving wife, for being with me throughout. It would have been impossible without you beside me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

1.1 Background

In the current world, mobility is a service. The infrastructure to accommodate mobility as well as introducing innovations in the way people and goods move ask for significant resources to build and maintain. However, the limitation of funds forces policymakers and planners to select the best alternative among several. Traffic and ridership forecasts accommodate this selection process by driving the benefit-cost analysis. Transportation engineers and planners use these forecasts as the justification of a project and a measure of scale. For example- if a road gets built or a new technology introduced, the change in the travel pattern or travel time will form the base of selection and hierarchy of projects. The number of people using the facility will also directly influence the dimensions of the project; for example, the number of lanes to be constructed, or the number of vehicles to be deployed, or even the toll rate.

The Fixing America'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 2016). Traffic forecasts are used, in part, to decide how these public dollars are invested, through environmental studies, capital cost estimations and benefit-cost analyses. Transportation infrastructure projects are typically huge investments, and the forecasts thus become important both in deciding whether to implement the projects at all and in prioritizing between projects selected for implementation. This "predict-and-provide" method means that inaccuracy in the traffic forecasts can have a great impact on the efficiency of the system design and the benefits accrued from the project against the cost. However, "*the greatest knowledge gap in US travel demand modeling is the unknown accuracy of US urban road traffic forecasts*" (Hartgen 2013). A relatively small set of empirical studies have examined non-tolled traffic forecasting accuracy in the United States. There is a need for research to expand the assessment and documentation of traffic forecasting experiences around the country to improve future modeling and forecasting applications, with the goal of ensuring that transportation funding dollars are being invested wisely.

1.2 Research Objectives and Problem Statement

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 assumptions inherent in 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. But undoubtedly, understanding the sources of error in traffic forecast is the first step towards refining the forecasting techniques. This has prompted quite a few ex-post evaluations of forecast accuracy in European countries as well as a few in the USA in recent years, although most of them fail to derive any substantial conclusion other than the measure of inaccuracy because of inadequate data (Nicolaisen and Driscoll 2014). In addition, the evaluation of forecast accuracy has primarily been concerned with toll-roads, where the inaccuracy has a greater bearing on investor confidence and project success. A relatively small set of empirical studies have examined non-tolled traffic forecasting accuracy in the United States. Here again, the lack of data makes rigorous statistical analysis difficult.

This study, funded by the National Cooperative Highway Research Program (NCHRP) project titled "Traffic Forecast Accuracy Assessment Research" (NCHRP 08-110), aims to fill the void of unknown traffic forecast accuracy in non-tolled roads. The objective is analyzing the accuracy and reliability of project-level traffic forecasts. This research attempts to answer these specific and complementary questions:

1. What is the distribution of forecast errors across the sample as a whole?
2. Can we detect bias in the forecasts?
3. Can we enumerate the sources of forecast error as hypothesized in previous research?

Taken together, answers to these questions will provide the means to describe the historic range of forecast errors that have been observed for certain types of projects. The analysis results will pave way to create an uncertainty envelope around the forecast traffic which can be a complementary technique to sensitivity analysis for risk assessment.

1.3 Research Approach

This study bases its analyses on the largest known collection of transportation projects compiled as part of the NCHRP funded project. The database contains information about 2300 projects with 16000 segments (or links) from six states in the USA and four European countries. The "Large-N Analysis" portion of the project, and the focus of this study, examines the overall amount and distribution of forecasting errors across different projects and agency characteristics, e.g. the methodology, forecast horizon, type of project, project area and functional class of the roadway, opening year and year forecast was produced to name a few. The study also identifies a potential effect of economic conditions, particularly unemployment rates on forecast accuracy.

The primary metric for evaluating the performance of the forecasts is the difference between the forecasted traffic and the earliest post-opening actual traffic. The inaccuracy is expressed as a percentage of the forecast volume because it is the value known at the time of forecast and thus the result can be used to estimate the uncertainty of the actual traffic.

1.4 Thesis Structure

As mentioned previously, this study is part of the NCHRP funded project *NCHRP 08-110: Traffic Forecast Accuracy Assessment Research*. The literature review to identify the gaps in knowledge in forecast accuracy and to establish the analysis procedure of the project itself as well as the analysis and their interpretation has been co-written by this author as part of the Large-N Analysis phase of the parent project. The chapters (the language and the organization) have been reproduced here with the Principle Investigator's permission from the main project documents, especially the Interim Report of the project and the Technical Report which await formal publication.

The thesis is organized in the following chapters:

Chapter 1 introduces the basic premise of the research- the research goals and objective as well as a brief overview of the approach. A brief history of research into forecast accuracy is reviewed in Chapter 2. This chapter also presents the existing systematic review programs for forecast performance evaluation. The goal of Chapter 2 is to establish the research goals by identifying the gaps in knowledge.

Chapter 3 establishes the procedure of analysis. It discusses the database that is used for analysis and identifies potential explanatory variables to base the analysis on. The second part of this chapter reviews the methods employed in existing researches and presents the analysis procedure of this study.

The analysis results are presented in Chapter 4. The distribution of forecast errors on the explanatory variables identified in Chapter 3 are documented along with the interpretations.

The final chapter of this document, Chapter 5, summarizes the findings from the research and identifies the limitations of the study and provides directives for future work stemming from this research.

CHAPTER 2. ASSESSMENT OF FORECAST ACCURACY- A LITERATURE REVIEW

2.1 Introduction

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. This chapter summarizes previous research works and what can be learned from them for the current study. It begins by reviewing a history of distinct forecast evaluation research studies demonstrating how the few efforts have been concentrated mostly on tolled roads and transit. Several existing systematic review programs for assessing forecast accuracy are reviewed next. This is followed by an examination of the best evidence on the accuracy of travel forecasts, summarized from a meta-analysis by Nicolaisen and Driscoll (2014). This review is aimed at serving the following two purposes:

1. To identify past works into forecast accuracy to ascertain the gaps in knowledge.
2. To establish the research goals and a base to compare the potential results to previous work.

This chapter borrows heavily from the NCHRP 08-110: *Traffic Forecast Accuracy Assessment Research Project* report, for which this literature review was specifically written.

2.2 A History of Forecast Evaluations

The decisions in public policy-making most often hinges on an apparent scientific evidence presented by the forecast of benefits and costs. This presents an ethical dilemma regarding the purpose of the forecast- whether to justify an already decided action or to evaluate several alternatives to choose the best one. In "*Ethics and Advocacy in Forecasting for Public Policy Decisions*", Martin Wachs explores these predicaments concerning traffic forecasts for large and costly infrastructure projects (Wachs 1990). The technical complexity associated with traffic forecasts is often misleading, he says, and forecasts are the "*elaboration of relatively simple assumptions about the future*". These assumptions can be tailored to fit a narrative; forecasts, after all, have a political use. He raises the question of deliberation: how much of the inaccuracy in forecasts are optimism bias and how much are deliberate.

The impact of the core assumptions on the accuracy, or the lack of, of forecasts is in support of William Asher's examination of forecasts in five areas: population, the economy (current dollar and real GNP), energy (electricity, petroleum consumption), transportation and technology (Ascher 1979). He found improvements in forecasting method a secondary precursor to achieving a higher degree of accuracy. According to Asher, failing to capture the reality of the future context leave little to the methodology. He also found that the more distant the forecast target date is, the less accurate becomes the forecast. He further identified systematic biases associated with the institutional sites of forecasts.

One other characteristic of forecasts is it's un-verifiability until the action has already been taken (Wachs 1990). In the context of transportation projects, evaluation of the accuracy has been done for several decades now, both as part of formal review process and independent researches. Table 1 summarizes key aspects of previous studies evaluating forecast accuracy, providing a survey of the history of forecast evaluations.

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). The BART was first of its kind in the United States- a regional rail system in an auto-centric metropolitan. The success of the project was hoped to have directed future forays into similar solutions to urban congestions. Webber compared the actual daily usage of the system as well as the effect on auto-ridership to the predicted. Early evidence suggested it being a "white-elephant": Webber's analysis found significant deviation of the actual scenario from the forecast. The total patronage of the system (average weekday trips) at 1976 was about half of what was predicted for 1975.

Table 1: Summary of existing studies

	Webber (BART)	Kahneman & Tversky	Flyvbjerg+UK Transport	Bain	NCHRP 364	TMIP Reports	FTA Capital Investment Grant Program	Minnesota, Ohio and Wisconsin Studies
Description of Issues/Challenges								
Institutional or Political constraints and influence			✓	✓				
Within Forecasting practice			✓	✓	✓			
Lack of archival practice					✓		✓	
Optimism Bias		✓	✓	✓	✓		✓	
Analysis of Predicted vs Actual Outcomes								
Demand forecasts	✓		✓	✓	✓		✓	✓
Project Benefits	✓		✓					
Project costs	✓		✓				✓	
Forecast assumptions/ exogenous forecasts	✓		✓		✓		✓	✓
Empirical assessment of new analytical methods					✓			
Suggested Changes to Methods/Practices								
Reference classes		✓	✓	✓				
Improved communication/reflection of uncertainty and/or risk		✓	✓	✓		✓	✓	
Verification of assumptions/exogenous forecasts						✓		
Identification of uncertainties and/or risks						✓		
Avoiding misapplication of model						✓		
Improved data reliability						✓	✓	
Improved model validation methods/practices						✓	✓	
Produce forecasts by independent parties		✓	✓	✓		✓		
Implemented Changes to Methods/Practices								
Reference classes		✓	✓	✓				

Improved communication/reflection of uncertainty and/or risk		✓	✓	✓			✓	
Identification of uncertainties and/or risks							✓	
Avoiding misapplication of model							✓	
Improved data reliability							✓	
New analytical methods							✓	
Improved model validation methods/practices							✓	
Produce forecasts by independent parties							✓	
Re-occurring reviews of predicted vs actual outcomes		✓					✓	
Analyzed Transportation Modes								
Public non-tolled roadways		✓				✓		✓
Tolled roadways		✓	✓	✓		✓		
Public transportation	✓		✓			✓	✓	
Outside transportation realm		✓			`			

Similar, but smaller scale, comparisons were made on other projects in the 1980s. A British study in 1981 examined the forecasts of 44 projects constructed between 1962-1971 (MacKinder and Evans 1981). The authors found no evidence that more recent or sophisticated modeling methods produced more accurate forecasts than earlier or more straightforward methods. In North America, the United States Department of Transportation produced a report in 1989 that examined the accuracy of 10 major transit investments funded by the federal government. This report (Pickrell 1989) concluded that most projects under-achieved their projected ridership, while simultaneously accruing capital and operating costs larger than expected. While the Pickrell Report and several other accuracy evaluations are focused on transit projects, the resulting criticism often extends to travel forecasting in general.

Similar to the analysis on BART, Dr. Kain looked into the Dallas Area Rapid Transit (DART) in 1990 (Kain 1990). He found that DART made extensive use of “clearly 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. 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 first examination into the reasons of travel forecast inaccuracy was an examination of the psychological biases in decision making under uncertainty in 1977. Kahneman and Tversky (1977) proposed the concept of the “inside view”, where intimate involvement with a project’s details during its planning and development phases leads to systemic over-estimates of its benefits and under-estimates of its costs. This was the first recognition of a systematic flaw in planning that is called “optimism bias” in today’s literature. The authors suggested the use of reference classes to correct these biases. 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.

The number of forecasting accuracy assessments have increased since the year 2000, although most of them have been 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. As an evidence to this, the Australia Government (2012) cited ‘‘*inaccurate and over-optimistic*’’ traffic forecasts as a threat to investor confidence. As Bain himself 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*” (Bain 2009). Three lawsuits are now underway that challenge the forecasts for toll road traffic that 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[1].

Bent Flyvbjerg released his seminal work on forecasts for multiple modes in 2005 (Flyvbjerg, Holm, and Buhl 2005). The article noted that demand forecasts were

[1] https://www.enr.com/articles/43707-arup-settles-17b-australia-toll-road-revenue-forecast-suit

generally inaccurate and not becoming more accurate over time. The conclusions were based on over 210 transportation projects (27 rail projects, 183 road) from across the world. The authors found that rail passenger forecasts are less accurate and more inflated than road vehicle forecasts at a very high level of statistical significance. This is not to say that road projects are more accurate, however, as the researchers found at least 25% of the projects go beyond the ±40% error range (difference between the actual traffic and forecasted traffic), 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 also identified potential causes for this inaccuracy, including 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. Flyvbjerg suggested developing and applying reference classes to projects with large uncertainties to get more accurate forecasts.

From 2002-2005, Standard & Poor's publicly released annual reports on the accuracy of toll road, bridge and tunnel projects worldwide. The 2005 report (Bain and Polakovic 2005), the most recent report available publicly, analyzed 104 toll road projects. They found that the demand forecasts for those projects were optimistically biased, and this bias persisted into the first five years of operation. They also found that variability of truck forecasts was much higher than lighter vehicles. The authors noted that their sample "*undoubtedly reflects an over-representation of toll facilities with higher credit quality*" and that actual demand accuracy for these types of projects is probably lower than documented in their report. The factors the researchers identified as drivers behind these errors 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).

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%. This observation is a stark contrast from other international studies, where the researchers have found a general over-estimation of traffic at a higher degree of inaccuracy. They attributed the standard organizational framework of a national toll forecasting system with "*little or no incentives to exaggerate the forecast*" as a factor.

Li and Hensher (2010) evaluated the accuracy of toll road traffic forecast in the Australian toll roads and found a general over-prediction of traffic. 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).

The National Cooperative Highway Research Program (NCHRP) released a synthesis on estimated toll road demand and revenue in 2006 (Kriger, Shiu, and Naylor 2006). This study reported the accuracy of 26 toll road revenue forecasts, finding that forecast accuracy does not improve over time. It noted that "*many of the problems that had been identified with the performance of traffic and revenue forecasts were related to the application of the model, less so to methods and algorithms*". More specifically, this finding is related to the assumptions needed to operationalize the models and not to the data or methods. It recommended analyzing the forecasting inputs and exogenous forecasts, and the improved treatment of uncertainties and risks.

Similar to the accuracy of toll road traffic forecasts, transit ridership forecasts have also attracted attention over the years. The BART and DART analysis (Webber 1976; Kain 1990) are examples of researches into this aspect. In more recent times, the Federal Transit Administration (FTA) has conducted two studies analyzing the predicted and actual outcomes of large-scale federally funded transit projects: one in 2003 (U.S. Department of Transportation: Federal Transit Administration 2003) and another in 2007 (Federal Transit Administration and Vanasse Hangen Brustlin 2008). 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 the projects New Starts built in the United States through 2011. The forecasts were incorporated into the Transit Forecasting Accuracy Database (TFAD). The database contained 65 large-scale transit infrastructure projects from around the country. The research found that transit project assumptions have historical bias towards over-forecasting ridership. Using this data, Schmitt statistically identified 3 reference classes for transit forecasting. The research also investigated three commonly held beliefs regarding forecasting accuracy:

1. More recent projects are more accurate than older ones (i.e., we are getting more accurate as tools become more advanced),
2. Forecasts are more accurate in later stages of project development than in earlier stages (i.e., the more we know about the details of a project the more accurately we can forecast demand), and
3. Forecasts of smaller changes to the transit system are more accurate than larger changes (i.e., smaller changes are easier to predict than larger changes).

It found that only the first commonly held belief had merit. Transit forecasts, on average, are biased but have been – slowly and non-monotonically – becoming more accurate over time. It is important to note, though, that this research has been focused on transit.

Compared to the analysis of accuracy for toll roads and transit projects, studies into non-tolled roadways are very few. Buck and Sillence (2014) demonstrated the value of using travel demand models in Wisconsin to improve traffic forecast accuracy and provided a framework for future accuracy studies. They evaluated 131 forecasts and determined the mean absolute difference between the forecasted and actual traffic to be 16%. On a much smaller scale, 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%. This result however is to be taken with a grain of salt, since they 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. They did not find any systematic problems with erroneous forecasts. The presentation also described an automated forecasting tool for “low risk” projects that relies on trend lines of historical traffic counts and adjustments following procedures outlined in NCHRP Report 255 (Pedersen and Samdahl 1982) and updated in NCHRP Report 765 (CDM Smith et al. 2014).

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.

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 (Flyvbjerg, Holm, and Buhl 2005; Buck and Sillence 2014; Parthasarathi and Levinson 2010). 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).

2.3 Existing Systematic Review Programs

Although individual studies analyzing the accuracy of travel forecasts are becoming more and more prevalent today, programs of forecast reviews are still rare. There are only three well-known re-occurring programs dedicated to reviewing predicted and actual outcomes already in practice.

The UK's Highways England in the Department for Transport, through their Post-Opening Project Evaluation (POPE) program (Highways England 2015), is the only known regular analytical review of non-tolled roadway forecasts in North America and Europe. It is by far the most impressive review of roadway forecasts. Highways England conducts a regular review of roadway forecasts, assessing the accuracy of demand, costs, accident, and travel time benefit forecasts. Over the past 11 years, the Highways England has reviewed smaller roadway projects (i.e., less than 10M British pounds). The Highways England also reviews large projects (i.e., greater than 10M British pounds) one and five years after each project's opening. A meta-analysis across all recent large projects occurs every two years.

The FTA's Capital Investment Grant program, commonly known as the "New Starts" program, requires Before and After Studies for every major project funded through the program (Federal Transit Administration 2016). Project sponsors are directed to archive the predictions and details supporting the predictions at two planning stages and at the federal funding decision stage. Approximately two years after project opening, project sponsors are required to gather information about the actual outcomes of five major aspects of the project: physical scope, capital cost, transit service levels, operating and maintenance costs and ridership. Project sponsors analyze the predictions and actual outcomes, and prepare a report summarizing the differences between the predictions and actual outcomes, documenting the reasons for those differences, and highlighting lessons learned that would inform FTA or other project sponsors on how methodologies or circumstances helped or hindered the predictions. FTA's New Starts program allows project sponsors to enumerate the uncertainties inherent in their travel forecasts and provide information on how those uncertainties may impact the project forecast. FTA has presented the method of "build up" of uncertainties, with separate forecasts produced for individual sources of uncertainty, to help identify the key drivers of uncertainty from the travel model's perspective. Similar approaches could be considered for highway projects.

The National Oceanic and Atmospheric Administration's Hurricane Forecasting Improvement Program (HFIP) is the only program that combines forecast accuracy evaluation with improved analytical methods, public communication of forecast uncertainty and societal benefits (National Oceanic and Atmospheric Administration 2010). The HFIP's stated accuracy goals were hypothesized to require increased precision in data and analytical methods. The HFIP developed a process to justify and evaluate these investments by placing analytical methods into three streams:

1. Stream 1 consists of existing analytical methods and is used for official, real-time forecasts;
2. Stream 2 consists of advanced analytical methods that take advantage of increased computing power and increased data precision, but forecasts are made offline; and
3. Stream 1.5 consists of elements of Streams 1 and 2 that seem to hold the most promise, forecasts are made in real-time but are not official.

The same input data is fed to all three streams. Efforts that demonstrate increased accuracy and skill are elevated to Stream 1.5 and eventually Stream 1. In this way, empirically proven methods are implemented very quickly. In five years, the HFIP has demonstrated a 10% improvement in tropical storm track and intensity forecasts (Toepfer 2015).

The HFIP is the only known program that uses a forecast skill metric in addition to traditional accuracy metrics. Advanced analytical methods must not only be accurate, but also must provide better accuracy than simpler and more inexpensive methods. In this way, analytical methods proven to be better than simpler (termed "naïve") methods are recommended for immediate implementation. Shortfalls in accuracy and skill are noted and used to prioritize future research efforts.

The HFIP directly tied improvement goals in forecast accuracy to societal benefits. "Forecasts of higher accuracy and greater reliability are expected to lead to higher user confidence and improved public response, resulting in savings of life and property" (National Oceanic and Atmospheric Administration 2010). As the first years of the program produced many successes, the accuracy goals were increased to eventually provide residents a reliable 7 days' advance warning of an impending storm. The estimated benefit of avoiding an unnecessary evacuation is $1,000 per person, and has been estimated to $225-380 million for larger storms (Toepfer 2015). In this way the HFIP sponsors can justify the cost of implementing more complex and expensive methods.

2.4 Summary of Existing Outcomes

Nicolaisen and Driscoll (2014) provided a recent meta-analysis of the demand forecast accuracy literature. That meta-analysis is not repeated here, but it is summarized to provide an existing baseline estimate of expected forecast accuracy.

Their analysis considers 12 studies that that have a sizable database of completed road and/or rail projects, that that provide distributions based on those projects, and that specify the sources of information considered. Table 2 shows the studies included, and Table 3 shows a summary of the results included. Both tables are reported directly from their paper.

Table 2: Summary of studies included in Nicolaisen and Driscoll (2014) meta-analysis

Prior studies of demand forecast inaccuracy included for review in the present paper

Author(s)	Projects opened	Area	Project
Mackinder and Evans (1981)	1970s	UK	Road
National Audit Office (NAO, 1988)	1980s	UK	Road
Pickrell (1990)	1980s	USA	Rail
Flyvbjerg, Holm, and Buhl (2006)	1970s–1990s	Global	Road + rail
Department of Transportation (DoT, 2007)	1990s	USA	Rail
DoT (2008)	2000s	USA	Rail
Bain (2009)	N/A	Global	Road
Button, Doh, Hardy, Yuan, and Xin (2010)	1970s–2000s	USA	Rail
Parthasarathi and Levinson (2010)	1960s–2000s	Minnesota	Road
Highways Agency (HA, 2011)	2000s	UK	Road
Welde and Odeck (2011)	2000s	Norway	Road
Nicolaisen (2012)	1970s–2010s	Scandinavia + UK	Road + rail

Table 3: Summary of results included in Nicolaisen and Driscoll (2014) meta-analysis

forecast inaccuracy

Author(s)	Sample[a]	Mean	Standard deviation
Mackinder and Evans (1981)	Road: 44	−7%[b]	N/A
NAO (1988)	Road: 128	+8%	43
Pickrell (1990)	Rail: 9	−65%	17
Flyvbjerg et al. (2006)	Road: 183	+10%	44
	Rail: 27	−40%	52
DoT (2007)	Rail: 19	−37%	31
DoT (2008)	Rail: 18	−16%	59
Bain (2009)	Toll: 104	−23%	26
Button et al. (2010)	Rail: 44[c]	−21%	58
Parthasarathi and Levinson (2010)	Road: 108	+6%	41
HA (2011)	Road: 62	+3	21
Welde and Odeck (2011)	Toll: 25	−3%	22
	Road: 25	+19%	21
Nicolaisen (2012)	Road: 146	+11%	35
	Rail: 31	−18%	33

[a]The three sample categories refer to rail projects (light, heavy, metro, etc.), road projects (highway, bridge, tunnel, etc.) and toll projects (same as road projects but with direct user charges).
[b]This value is for screen lines rather than individual stretches due to lack of data for the latter in this study.
[c]A number of fixed-guideway bus rapid transit (BRT) systems are also included in this study. However, no distinction between BRT and rail has been made in the present review as it has not been possible to obtain separate data for these two categories.

Their main finding is that the observed inaccuracy of forecasts varies based on the type of project:

1. For rail projects, the mean inaccuracy is negative, meaning that actual demand is less than the demand that was predicted. The general range is that actual demand is 16-44% less than forecast demand.

2. For toll road projects, the mean inaccuracy is also negative, indicating that actual demand is less than forecast.

3. For un-tolled road projects, the mean inaccuracy is positive, with most results showing 3-11% more traffic in reality than was forecast.

They also note that for all types of projects, there is considerable variation in the results, regardless of the mean. It should be noted, that there are limited studies available here, particularly of un-tolled roads in the United States, so these results should be considered with a degree of caution. Nonetheless, it is interesting to note the difference in direction for un-tolled road projects relative to rail and toll road projects, with the forecasts predicting too little demand for un-tolled roads, and too much demand for rail and toll roads. One can hypothesize possible explanations for this difference. Some possible explanations may be:

- There could be a methodological difference such that transit 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 Flyvbjerg (2007) 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.

While it is easy to speculate on the possible sources of errors, it is difficult to know for certain what the issue is. As Nicolaisen and Driscoll note: "*The studies that make the greatest effort to address this aspect are rarely able to provide more than rough indications of causal mechanisms.*" They go on to point out that a key challenge is the lack of the necessary data to conduct such studies, in particular, the infrequent availability of archived forecasts. Specifically, they point out: "*The lack of availability for necessary data items is a general problem and probably the biggest limitation to advances in the field*."

2.5 Gaps in Knowledge

The research reviewed here provide a starting point for understanding existing evidence on forecast accuracy, as well as a strong foundation of how to approach such studies and what factors may contribute to inaccuracy. We can identify a few common patterns and limitations in the studies that have been reviewed:

- Most of the past studies have focused on toll roads (Bain 2009; Odeck and Welde 2017; Kriger, Shiu, and Naylor 2006)) and transit projects (U.S. Department of Transportation: Federal Transit Administration 2003; Schmitt 2016; Voulgaris 2017).
- Accuracy of non-tolled roadway forecasts have not garnered much attention. In the US, accuracy assessment studies have been very few (Miller et al. 2016; Buck and Sillence 2014; Parthasarathi and Levinson 2010; Giaimo and Byram 2013).
- Even for the existing studies into this topic, the sample sizes are too little to arrive at any statistically satisfying conclusions regarding the factors behind the inaccuracy.
- In addition, the studies reviewed here also note the lack of data items available in their research.

It is from this point that this research begins—limited studies on un-tolled roads in the US, little information on the sources of forecast errors, and a general lack of data to conduct such studies. Revisiting the research questions outlined in the first chapter, we can set the more specific objectives for this study:

- Establish a database with enough data on forecasts and traffic to get statistically significant results.
- Establish an analysis procedure to identify biases.
- Analyze the forecast accuracy over several explanatory variables to identify the possible sources of error.

CHAPTER 3. PROCEDURE OF ANALYSIS: THE DATA AND THE METHODOLOGY

3.1 Introduction

As delineated in the previous section, the progress in assessing the accuracy of traffic forecast is hampered by the lack of data. Rigorous studies do exist, but the sample size is either too small or are unevenly focused which doesn't allow for systematic statistical analysis. A key reason for this data deficiency is that it takes proactive planning to accumulate the data necessary for retrospective analysis. This data is often lost, as forecast preservation and archival procedures are uncommon in practice, and long project development cycles and staff attrition make recovering this information very challenging. The NCHRP 08-110 project, on which this research is based, starts off by accumulating a database from various agencies across the United States. The database currently contains forecast information on about 16,600 segments or links across 2300 different projects in the six participating states as well as four European countries. In the first section of this chapter, the structure of the database is reviewed to identify the potential explanatory variables for forecast accuracy. Next, the method for analysis is presented along with a brief review of the existing works.

3.2 Available Data and Key Challenges

This analysis uses the database compiled as part of the NCHRP 08-110 project. The database contains traffic forecast and actual traffic information for road projects in several states. The sources are the Department of Transportation (DOT) maintained databases, Equivalent Single Axle Loading (ESAL) forecast reports, project 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.

3.2.1 Data Characteristics

Data from six states (Florida, Massachusetts, Michigan, Minnesota, Ohio and Wisconsin) have made up the database. It also contains a separate database compiled by Nicolaisen (2012) which has data from four European countries (Denmark, Norway, Sweden and the United Kingdom). A short summary of the available information, with the State names replaced by Agency Code to protect anonymity, is presented in Table 4:

Table 4: Summary of Available Data

Agency	All Projects		Opened Projects	
	Number of Segments	**Number of Unique Projects**	**Number of Segments**	**Number of Unique Projects**
Agency A	1123	385	425	381
Agency B	12	1	12	1
Agency C	38	7	5	3
Agency D	2176	103	1292	99
Agency E	12413	1863	1242	562
Agency F	463	132	463	132
Agency G	472	120	472	113
Total Segments	16697	2611	3911	1291

In total, the database contains reports for 2,611 unique projects, with 16,697 segments associated with those projects. A segment is a different portion of roadway for which a forecast is provided. For example, forecasts for an interchange improvement project may contain segment-level estimates for both directions of the freeway, for both directions of the crossing arterial, and for each of the ramps. Some of these projects have not yet opened; some of the segments do not have actual traffic count data associated with them, and others do not pass the quality control checks for inclusion in the statistical analysis (the filtering process is described later in Section 3.4). While all records are retained for future use, the analysis is based on a filtered subset of 1,291 projects and 3,911 segments.

The different sources of datasets naturally lead to inconsistency in the way the data are stored. Key fields that may correlate with forecast inaccuracy as identified in our literature review are missing in few, and they are all provided in different formats- scanned reports, excel tables, database from previous studies. Actual traffic counts, when absent in the documents provided by the agencies, were collected from different sources- historical count archives and count maps.

The opening year of the projects in the database varies from 1970 to 2017, with about 90% of the projects opening in year 2003 or later. While the exact nature and scale of the project isn't always known, inspection reveals that the older projects are more likely to be major infrastructure projects, and the newer projects are more likely to be routine work for the DOT, e.g. resurfacing works on existing roadway. For example, almost half of the projects are design forecasts for repaving. Such differences are driven largely by data availability. Some state agencies have recently begun tracking the forecasts, and the records to do so rarely go back more than 10-15 years. The older projects are derived from someone going back to study and enter paper reports or scans of paper reports, with the availability of documentation and the interest in spending the effort to examine higher for bigger projects. Thus, it is not a random sample of projects, and there are notable differences not only in the methods used across agencies, but also in the mix of projects included in the database. This is an important limitation that readers should bear in mind as they understand and interpret the results.

The agencies have about two-thirds of the data items filled in, but they are not the same two-thirds every time. The absence of data fields required us to make assumptions specific to the states and the data characteristics. For example: in the Minnesota data, not much information is available in the reports regarding how the forecast was made. For forecasts that have been done before 1980, it is assumed that the forecast was made using traffic count trend analysis. In several other DOTs case, while actual counts were given on the same roadway, there was no mention of when the project was completed. Missing key information like Project Type/Type of Improvement, Roadway Facility Functional Class, Forecast Methodology etc. were more common.

The most important assumption has been made for the actual traffic count. For a correct measurement of forecast inaccuracy, the forecasted traffic and the actual traffic need to be on the same year after the project has completed. But most of the sources don't indicate if the actual count was taken on or after the year project was actually open for use. The Wisconsin and Minnesota datasets come from two published researches on assessing forecast accuracy: Buck and Sillence (2014) for Wisconsin and Parthasarathi and Levinson (2010) for Minnesota. The Florida D-4 data were obtained from a published study as well (Traffic Forecasting Sensitivity Analysis, 2015) which compares the actual count in the forecasted year with the forecasted traffic. For these datasets we can assume that the Actual Traffic Count listed are taken after the project has been completed. As for the Ohio dataset, the actual year of completion of the projects were given only for a few records. For others, there was no indication whether the counts are taken after the project has opened or not. Similarly, Florida District 5 datasets were compiled from ESAL reports. Here again we don't have any indication of the actual opening year of the projects. In such cases, we have taken a traffic count a couple of years after the project is forecasted to open and scaled the forecast values up to that year. The assumptions for filtering and cleaning the data is described in Section 3.4.

3.2.2 Database Structure

The primary fields on the Forecast Database can be classified into three types:

1. Project Information
2. Forecast Information and
3. Actual Traffic Count Information

Project Information table has all the information specific to the project characteristics. This includes Project/Report ID unique to a project, Project Description, Year when the project/report was completed, type of project, City or Location where project took place, State, Construction cost, etc. Forecast Information includes the data related to the traffic forecast: the forecast itself along with who made the forecast, at and for what year. It also includes the type of forecast year (opening, mid-design or design year), the methodology used to forecast, whether any post-processing been done or not and similar information. Information regarding the actual traffic includes the actual traffic volume in a particular segment, year of observation and project opening year. The key fields in the database is given in Table 5.

Table 5: Key Fields in NCHRP 08-110 Database

Name	Description
Brief Description	Brief written description of the project
Project Year	Year of the project or Construction Year or the Year the Forecast Report was produced
Length	Project Length in miles
Functional Class	Type of facility (Interstate, Ramp, Major/Minor Arterial etc.)
Improvement Type	Type of project (Resurfacing, Adding lanes, New construction etc.)
Area Type Functional Class	Area type where the facility lies (Rural, Urban etc.)
Construction Cost	Project construction cost
State	State code.
Internal Project ID	Project ID or Report ID or Request ID
County	County in which the facility lies
Toll Type	What kind of tolls are applied on the facility (No tolls, Static, Dynamic etc.)
Year of Observation	Year the actual traffic count was collected
Count	Actual Traffic Count
Count Units	Unit of traffic count (AADT, AAWT).
Station Identifier	Count station ID or other identifiers for count station.
Traffic Forecast	Forecasted Traffic volume.
Forecast Units	Units used to forecast traffic (AADT, AAWT).
Forecast Year	Year of forecast.
Forecast Year Type	Period of forecast like opening, mid-design or design period.
Year Forecast Produced	The year the forecast was produced/generated.
Forecasting Agency	Organization which was responsible for this forecast.
Forecast Methodology	Method used to forecast traffic (Traffic Count Trend, Regional Travel Demand Model, Project Specific Model etc.)
Post Processing Methodology	Any post processing or alternative methodology used.

Post Processing Explanation	Explanation, as warranted, in case post processing methodology is used.
Segment Description	Description of the segment for which this forecast was done.

3.2.3 Decision Variables

Based on the nature of the NCHRP 08-110 database, we selected some variables that can explain the bias in forecasts. These variables are: the type of Project, the methodology used, roadway type, area type and the forecast horizon (difference between the year forecast produced and the year of opening).

Project Types are coded into the database as Improvement Type. Along with unknown improvement types, the improvement types are categorized into 12 types, which are consolidated into Projects on Existing Roadway, New Construction Project and Unknown Project Type (Table 6).

Table 6: Description of Project Types in the NCHRP Database

ID in Database	Improvement Type	Unified Improvement Type
1	Resurfacing/Replacement/no minor improvements	Project on Existing Roadway
2	In existing facility, add intersection capacity	
3	In existing facility, add mainline/mid-block capacity in general purpose lane(s)	
4	In existing facility, add new dedicated lane(s)	
5	In existing facility, add new managed lane(s)	
6	In existing facility, add new reversible lane(s)	
7	New general-purpose lane(s) facility	New Construction Project
8	New dedicated lane(s) facility	
9	New managed lane(s) facility	
10	New reversible lane(s) facility	
11	Other New Facility	
12	Unknown Improvement	Unknown Project Type

The Functional Class column in the database are coded according to the FHWA specified functional classification. For a few datasets, the functional classes of the roadway were provided in an older format, which were then converted into the new format (Table 7).

Table 7: Description of Functional Class in the NCHRP Database

ID in Database	Functional Class
1	Interstate or Limited-access facility
2	Ramp
3	Principal Arterial
4	Minor Arterial
5	Major Collector
6	Minor Collector
7	Local
8	Unknown Functional Class

The area type or the County where the facility lies is mainly coded in four categories: Rural, Mostly Rural, Urban and Unknown area types (Table 8). The definition of these categories is consistent with the 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%).

Table 8: Description of Area Type in the NCHRP Database

ID in Database	Area Type
1	Rural
2	Mostly Rural
3	Urban
4	Unknown Area Type

Forecast Methodology were identified from the project reports or the datasets given by the State DOTs. For example, for the Florida D-4 dataset, the methodology was derived from the Method column and then were reassigned into the NCHRP methodology (Table 9). For most of the database where the methodology is not clearly described, several assumptions have been made (see previous section) to sort them by the NCHRP codes.

Table 9: Description of Forecast Methodology in the NCHRP Database

ID in Database	Forecast Methodology	Explanation
1	Traffic Count Trend	Compound and Linear Growth Rate, Linear Interpolation, Regression Models etc. using Historical AADT or traffic count on a specific count station.
2	Population Growth Rates	Forecasts based on Socio-Economic data, population forecasts on TAZ or project catchment area.
3	Project-Specific Travel Model	Travel Demand Model created specifically for a project.
4	Regional Travel Demand Model	Travel Demand Model for a region, e.g. Central Florida Regional Planning Model (CFRPM), Florida Standard Urban Transportation Model Structure (FSUTMS) etc.
5	Professional Judgement	Usually a combination of traffic count trend and Travel Demand Model volume.
6	Unknown Methodology	No record of methodology used.

Several assumptions have also been made to code the Forecasting Agency in the NCHRP format (Table 10). For example, for Florida District 4, Minnesota and Wisconsin projects, the agency has been assumed to be State DOT employees or members. Consultants under contract with State DOTs (like Florida District 5 projects) were categorized separately.

Table 10: Description of Forecasting Agency in the NCHRP Database

ID in Database	Forecast Agency
1	State DOT
2	Metropolitan Planning Organization
3	City/County agency
4	Other public agency
5	Consultant

Two other variables we have included in our final estimation dataset are the Unemployment Rates in the year the forecast was produced and in the opening year. These data are not provided by the agencies themselves and that is why they are absent in the main database. We have collected the Unemployment Rate at a state level for the US projects from the Bureau of Labor Statistics and at the national level for the European projects from the World Bank.

3.3 Methodology

This study uses Large-N analysis to measure the amount and distribution of forecast errors, including those segmented by variables such as project type and various risk factors. Large-N studies consider a larger sample of projects in less depth. Flyvbjerg (2005) extols the virtues of Large-N studies as the necessary means of coming to general conclusions. Often, Large-N studies include a statistical analysis of the error and bias observed in forecasts compared to actual data. Flyvbjerg et al. (2006) considered a Large N analysis of 183 road and 27 rail projects, and Standard and Poor's conducted a Large N analysis with a sample of 150 toll road forecasts (Bain and Plantagie 2004). Other examples of Large-N studies are the Minnesota, Wisconsin and Ohio analyses (Parthasarathi and Levinson 2010; Buck and Sillence 2014; Giaimo and Byram 2013). This section presents a brief overview of the methodologies used in existing literature and explains the methodology used in current research.

3.3.1 Methodologies Used in Existing Literature

Briefly, the goal of Large-N analysis is to answer: How close were the forecasts to observed volumes? (Miller et al. 2016). In order to facilitate that, the researchers have generally looked at two sets of similar data: one during the base year and the other one in the forecast year. Several authors have evaluated the accuracy of project level traffic forecasts by comparing them with the actual traffic counts. A summary of existing research and the methodology used is given in Table 11.

Table 11: Summary of Existing Large-N Methodologies

Paper	**Research Data**	**Analysis Procedure**
Odeck and Welde (2017)	68 Norwegian Toll Road Projects implemented between 1975 and 2013.	Mean Percentage Error of actual traffic compared with forecast value, Examines bias and efficiency of estimates using econometric framework
Li and Hensher (2010)	14 Toll Roads in Australia	Mean Percentage error of actual vs forecast traffic. Ordinary Least Square Regression model and Random effects regression models with Percentage Error as dependent variable to ascertain the biases and dependencies.
Flyvbjerg et al. (2006)	183 projects around the world	Percentage Error with actual vs forecast traffic to measure inaccuracy.
Bain (2009)	104 international toll road, bridge, and tunnel case studies.	Ratio of Actual and forecasted traffic.
Miller, Anam, Amanin, and Matteo (2016)	39 studies from Virginia	Mean Absolute Percentage Error for each segment, Median Absolute Percentage Error for individual projects (both compared over the Observed Value).
Parthasarathi and Levinson (2010)	108 project reports obtained from Minnesota DOT.	Ratio of Actual and forecasted traffic

3.3.2 Evaluation Year

From the database and project reports, we see that traffic forecasts are usually done for three years:

1. Opening Year
2. Mid-Design or Interim Year (usually 10 years after Opening)
3. Design Year (usually 20 years from Opening)

The actual traffic counts are obtained from the DoT's count stations. For example, the Florida District 5 has detailed traffic counts from their count stations from

1972 to 2016. Matching the Count Stations with the traffic forecast report, we can get the actual traffic count for a year. Three calculations of errors can be performed:

1. Error in the opening year forecast
2. Error in Interim/Mid-Design years
3. Error in year in-between Opening and Mid-Design Year: In this case, the forecast traffic value can be interpolated.

The purpose of taking errors or difference in forecasted traffic for different years is to evaluate whether forecast performance improves over time. Li and Hensher (2010) report that all other factors remaining unchanged, the error in forecast reduces by 2.54% for every additional year since opening i.e. we see annual improvements, on average, in the accuracy of forecasts as we move away from the start date. This finding is supported by Vassallo and Baeza (2007) with the evidence that traffic forecasting effectiveness for Spanish toll roads tends to improve over time, in particular the research claimed that the average year-one error was -35.18%, -31.14% for the second year, and -27.06% for the third.

This research will focus on the evaluation of opening year forecasts for the practical reason that the interim and design years have not yet been reached for the vast majority of projects.

3.3.3 Definition of Error

One of the differences in methodologies in previous Large N studies is how they define errors. Miller et al. (2016), CDM Smith et al. (2014), and Tsai, Mulley, and Clifton (2014) define error as the Predicted Volume minus the Actual Volume such that a positive result is an over-prediction. Odeck and Welde (2017), Welde and Odeck (2011), and Flyvbjerg, Holm, and Buhl (2005) defined error the other way, such that a positive value represents under-prediction.

There are also two schools of thought when presenting the error as a percentage: over the actual traffic (Tsai, Mulley, and Clifton 2014; Miller et al. 2016) vs over the forecast traffic (Flyvbjerg, Holm, and Buhl 2005; Nicolaisen and Næss 2015; Odeck and Welde 2017). An advantage of the former is that the percentage is expressed in terms of a real quantity (observed traffic); an advantage of the latter is that when the forecast is made, uncertainty can be expressed in terms of the forecast value since the observed value is unknown (Miller et al. 2016). Beside these two methods, Bain (2009) and Parthasarathi and Levinson (2010) evaluated the forecast performance by taking the ratio of Actual and Forecast Traffic.

From the discussion above and the summary in Table 11, we see basically two schemes for evaluating forecast performance: as a percentage difference from forecast volume and as a ratio. Within those schemes, there is some disagreement as to whether the percentage difference should be taken over the observed count or over the forecast value, and as to the direction of the sign. For this we continue in the convention as described in (Odeck and Welde 2017) in expressing the percent difference as:

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

Where PDFi is the Percentage Difference from Forecast for project i. Negative values indicate that the actual outcome is lower than the forecast (over-prediction), and positive values indicate the actual outcome is higher than the forecast (under-prediction). The appeal of this expression is that it expresses the deviation as a function of the forecast, which is the value known at the time of forecast. The distribution of the Percent Difference over the dataset will be able to answer the systematic performance of traffic forecasts.

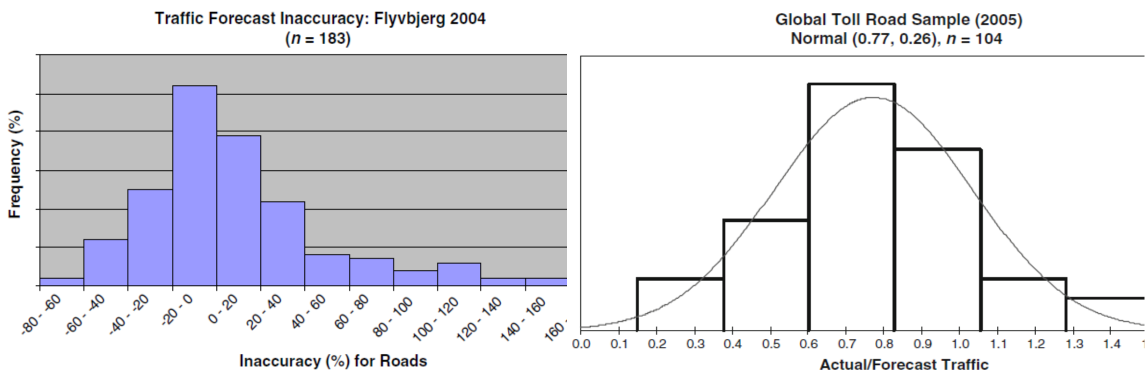
As for expressing the error over the dataset, the use of Mean PDF and Mean Absolute PDF have varied in different researches. Taking the mean of the absolute differences has been acknowledged to "*allow [researchers] to better understand the absolute size of inaccuracies across project*" (Odeck and Welde 2017) since positive and negative errors tend to offset each other in case of calculating the Mean Percent Difference. We continue in this tradition, and express:

$$Mean\ Absolute\ Percent\ Difference\ from\ Forecast\ (MAPDF) = \frac{1}{n} * \sum_{i=1}^{n} |PDF_i| \qquad Equation\ 2$$

Where n is the total number of projects.

3.3.4 Distribution of Percent Difference from Forecast

Researchers have presented the results of their Large-N studies mostly in histograms of Percentage Error, as shown in Figure 1. Bain (2009) further fitted the distribution in a distribution fitting software, which suggested a normal distribution with mean 0.77 and Standard Deviation 0.26. Goodness of fit was measured by Chi-squared statistics. To ascertain the significance of the statistics (biasedness), t-test was also performed.



Source: Flyvbjerg, Holm, and Buhl (2005) Source: Bain (2009a)

Figure 1: Example Histograms of Forecast Accuracy

This research reports distributions of the errors in terms of the percentage difference from forecast, PDF_i. The mean as reported in the distribution gives the central tendency of the dataset, with median as the 50th percentile value and standard deviation as the spread. For categorical variables, this research employs Violin Plots (Figure 2). Violin plots are like histograms and box plots in that they show an abstract representation of the probability distribution of the sample. Rather than showing counts of data points that fall into bins or order statistics, violin plots use kernel density estimation (KDE) to compute an empirical distribution of the sample. In this research, we used the 5th and 95th percentile values as inter-quartile range as depicted in Figure 2. The percentile values present the percentage of data-points that fall below. In effect, this range depicts the 90% probability range of percent difference from forecast for any categorical variable.

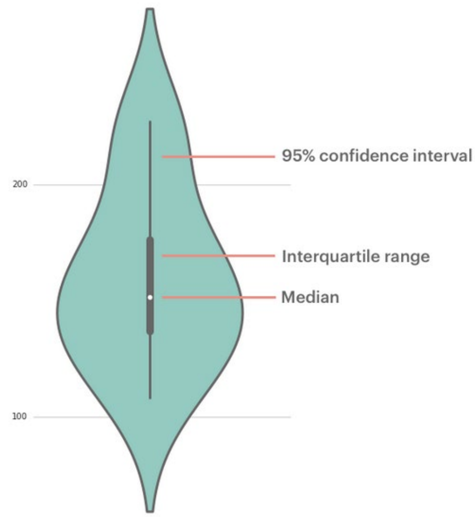


Figure 2: Anatomy of a Violin Plot

3.3.5 Level of Analysis- by Segment or by Project

While assessing the project forecast accuracy, one question arises: what constitutes an observation? A typical road project is usually divided into several links or segments within the project boundary. The links are usually on different alignments or carrying traffic to different directions. Analysis thus can be done on two levels:

1. Segment Level: assessing the accuracy of the forecast for each different segment or link.
2. Project Level: assessing the total accuracy of forecast for each individual project, identified by their Unique Internal Project ID.

The limitation of presenting accuracy metrics at a segment level is that the observations are not independent. Consider, for example, a project with three segments connected end-to-end. It is reasonable to expect that the forecast error on these segments is correlated—perhaps uniformly high or low. Whether we treat these as one combined observation or three independent observations, we would expect the average error to be roughly the same. There would be a difference, however, in the measured t-statistics, where the larger sample size from a segment level analysis could suggest significance where a project level analysis would not.

Project level analysis seems to be free of the correlation across observations described, but still the question remains on how to assess the accuracy for a project. In the Virginia Study (Miller et al. 2016) where each project consisted of links ranging from 1 to 2493 in number, the researchers took the Median Absolute Percent Error over the segments or links for individual projects and then used the Mean to express the level of accuracy. Nicolaisen (2012) measured the accuracy by taking the sum of forecast and actual traffic volumes on the segments in a project. Another method that can be used is taking the weighted traffic volume as described in Miller et al. (2016):

$$Weighted\ Traffic\ Volume = \frac{\sum_{i=1}^{n}(Volume\ on\ link\ i) * (Length\ of\ link\ i)}{\sum_{i=1}^{n}(Length\ of\ link\ i)}$$

$Equation$ 3

The issue with using the weighted traffic volume (forecasted and actual) is the absence of length data in most of the records. In addition, taking the total traffic as Nicolaisen (2012) will not be able to show the relation between forecast accuracy and project type by number of vehicles serviced. Taking these into consideration, in this study we measure the inaccuracy at the project-level using average traffic volumes, where each segment within a project is given equal weight.

We report the distribution of forecast errors both at a project level and a segment level. The results, presented later in next chapter, show that averaging to the project level appears to average out some of the errors observed at a segment level.

3.4 Calculating the Number of Lanes Required

One of the implications of inaccurate forecast is how it would influence project decisions. The Number of Lanes required for the roadway to operate at a certain Level of Service is a variable dependent on the anticipated traffic. Miller et al. (2016) in the Virginia Study explored a variant of this in the decision concerning the Level of Service (LOS). One of the projects (or studies as the research termed it as), had seen an LOS E instead of the target LOS of C because of forecast errors. The research identified two distinct factors that affect the impact of error on decision making:

1. The magnitude of the error and

2. The location of the error relative to the performance criterion.

Replicating the methodology employed in the Virginia Study in our analysis is problematic because of the absence of several critical information to calculate the LOS. The existing and forecasted number of lanes and the K-factor used was not specified for most of the projects and we would be dealing with a very small sample size. Besides, other factors influencing the LOS e.g. Lane Width, Traffic Composition, Grade and Speed were not coded into the database.

Another way to assess the impact of forecast error is to calculate the number of lanes required for a given traffic volume. Project traffic forecasts ultimately are used to determine how many lanes a corridor or project may require. Using the best available current year data, and projecting future values of Directional Design Hourly Volume (DDHV), Service Flow Rate for LOS I (SF_i)and Peak Hour Factor(PHF), the number of lanes can be estimated. Using the methodology described in Highway Capacity Manual-2010 (HCM 2010) to calculate the Service Flow Rate per lane for a required LOS and PHF, the number of lanes can be determined. According to it, the simplified equation for estimating the capacity of a roadway section is:

$$Capacity = Base\ Capacity * N * PHF * f_{HV} * f_p \qquad Equation\ 4$$

where N= Number of Lanes

PHF= Peak Hour Factor

f_{HV}=adjustment factor for heavy vehicles

f_p =adjustment factor for driver population

Rearranging the equation to determine the number of lanes for given traffic flow on a given direction, we get to:

$$N = \frac{Traffic\ Volume\ on\ a\ given\ direction}{Base\ Capacity * PHF * f_{HV} * f_p} \qquad Equation\ 5$$

The Traffic Volume on a Given Direction can be alternately named as Directional Design Hourly Volume, which can be determined using:

$$DDHV = AADT * Design\ Hour\ factor\ (K_{30}) * Directional\ Distribution\ factor\ (D_{30}) \quad Equation\ 6$$

The K-factors represent typical conditions found around the state for relatively free-flow conditions and are considered to represent typical traffic demand on similar roads. The magnitude of the K-factor is directly related to the variability of traffic over time. Rural and recreational travel routes which are subject to occasional extreme traffic volumes generally exhibit the highest K-factors. The millions of tourists traveling on Interstate highways during a holiday are typical examples of the effect of recreational travel periods. Urban highways, with their repeating pattern of home-to-work trips, generally show less variability and, thus, have lower K-factors. Similarly, the directional distribution factor, D_{30}, is based on the 200th Highest Hour Traffic Count Report. But the problem remains as to the availability of K_{30} and D_{30} information for projects. The Florida Department of Transport (FDOT) recommends values for the K and D-factor in case information on that is unavailable during project forecast. The following table is obtained from the Project Traffic Forecasting Handbook prepared by the FDOT (FDOT 2014).

Table 12: Recommended $\boldsymbol{K_{30}}$ and $\boldsymbol{D_{30}}$ factors for Traffic Forecasting

Road Type	$\boldsymbol{K_{30}}$			$\boldsymbol{D_{30}}$		
	Low	Average	High	Low	Average	High
Rural Freeway	9.6	11.8	14.6	52.3	54.8	57.3
Rural Arterial	9.4	11	15.6	51.1	58.1	49.6
Urban Freeway	9.4	9.7	10	50.4	55.8	61.2
Urban Arterial	9.2	10.2	11.5	50.8	57.9	67.1

HCM recommended range of values for selecting appropriate $K_{30}$and D_{30}factors for project forecast are also given in the following figures.

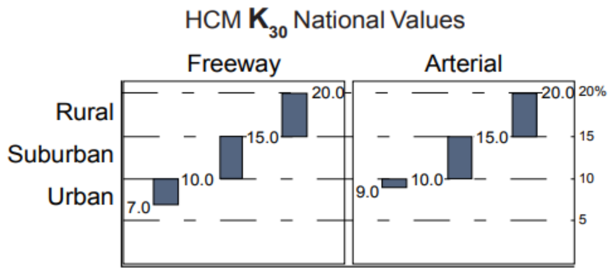


Figure 3: HCM Recommended K Factor Range

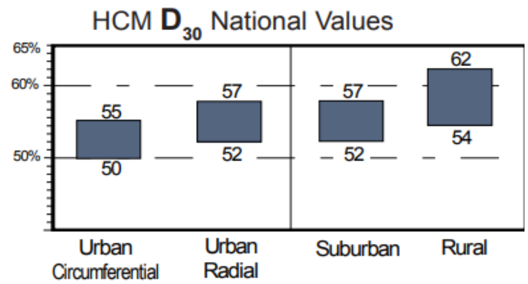


Figure 4: HCM Recommended D Factor Range

For a simple analysis, we chose the average values in each subsection as recommended by the FDOT.

The equations for determining the base capacity for the roadway types are also recommended in HCM 2010, which are presented in the . In the absence of information on Free Flow Speed, in our analysis we are assuming the maximum lane capacities by default.

Table 13: Equations to Determine Service Flow Rate or Maximum Capacity

Roadway Type	Equation
Freeway (Interstate)	1700+10*Free Flow Speed (FFS) up to 2400
Multilane Highway	1000+20FFS up to 2200
Rural 2-lane Highway	Up to 1600
Signal Controlled Facility	1900*green ratio

The Peak Hour Factors (PHF) are taken as the default values given in HCM 2000: 0.92 for Urban facilities and 0.88 for rural ones.

Assuming similar LOS for forecasted traffic and actual traffic and using Equations 10 and 11, we first calculated the number of lanes required for each case and then compared them with each other. Note, we used the upper bounds for the N values, as specified in HCM.

3.5 Data Cleaning and Filtering

As mentioned previously, our primary objective for analysis is to compare the forecasted traffic with the actual post-opening traffic. The NCHRP 08-110 Database presents challenges in the analysis due to the difference in record keeping practices of the contributing states (explained in Section 3.2.1). We arrived at a uniform scheme or algorithm to clean up the missing information and prepare the flat data for analysis.

First of all, we filtered out the records in the database that don't have any actual traffic count data and those which haven't been completed yet. The second filter may seem redundant, but in the database we have records of actual traffic count even though the project was forecasted to be completed at a later date. This discrepancy occurred mostly for projects on existing roadways that have traffic count stations on them which produce regular count data.

The second step was to select the appropriate actual traffic count for the records filtered out in the first step. This was necessary because in many cases traffic counts were collected on a regular basis on the same segments over several years. Selecting the earliest traffic count after project completion is often not obvious, because several state data don't mention actual project completion date. For such types of projects, we employed the following reasoning:

1. Categorize the segments by schedule risk: Based on the improvement types, we created low-risk and high-risk categories. The "resurfacing, slips, slides, safety improvements etc." projects that are usually completed on or within one year of the planned opening year are the low-risk ones. Complex projects like adding lanes to a roadway, new construction projects or increasing capacity are built within two to three years of the planned opening date and are therefore classified as High-Risk ones (Mark Byram, Personal Communication, April 3, 2018).
2. Create a one-year buffer for low-risk project and a two-year buffer for high-risk project and keep the first traffic count outside the buffer. For example, if a project to add lanes, a High-Risk project, has a forecast opening year of 2010, we would keep the first count available in year 2012 or later. We do this because we do not know if construction has been delayed from what was originally planned, and we want to avoid a situation where we evaluate a project against a traffic count taken before the project opened.

Next, we scaled the forecast to the year of the first post-opening count so that both data points are in the same year. We did this by linearly interpolating the forecast traffic between the forecast opening year and the design year, usually 20 years later. (The European projects are taken from Nicolaisen (2012) and have already been scaled to match the count year using a 1.5% annual growth rate. We maintain this logic for the European projects but do the interpolation between opening and design year for US projects.)

For project-level analysis, we took the average of the traffic volumes and measured the error statistics by comparing the average forecast and average actual traffic. Aggregating the counts and forecast across the segments/links was done by the unique identifier in the column "Internal Project ID". The variables for analysis were also aggregated by the same unique identifier, albeit with different measures for maintaining uniformity. Improvement Type, Area Type and Functional Class of a project were taken to be the same as the most prevalent one among the segments. For example, if most of the segments in a project are of Improvement Type 1 (resurfacing/ regular maintenance), the project is considered to be of Improvement Type 1. Forecast Methodology is the same across the segments for a project, so are Unemployment Rates and Years of Forecast and Observation. Mean of these values were taken for the project level analysis.

CHAPTER 4. ANALYSIS RESULTS

4.1 Introduction

This section presents the key findings from the Large N Analysis building on the methodology prescribed in Section 3.3. Reiterating the key points our analysis hinges upon:

- typical road projects are divided into one or more segments
- traffic volume is generally predicted for opening year, mid-design year (typically 10 years from opening) and design year (usually 20 years into the future)
- actual traffic volume to compare against the forecast volume are taken for the year after the project has been completed. For records in the database that don’t have project completion date, a buffer of at least 1 year has been created based on the type of project
- error is calculated as the difference between Actual Volume and Forecasted Volume and so that negative value means over-prediction and positive means under-prediction
- for aggregation, the Mean of the Absolute Percent Difference from Forecast (MAPDF) was used as the metric, since positive and negative values would neutralize each other in case the mean of the percent differences were taken. The distributions, however, were taken on the Percent Difference from Forecast (PDF).

Bearing these points in mind, the Large N analysis was done in two ways: by segments for the general distribution of the forecast errors and by project-level, for the effect of errors on an aggregated level. The variables to analyze are introduced in the first section and we move onto the results and interpretation in the next section.

4.2 Data and Variables: A Recap

As described Section 3.2, the NCHRP 08-110 database contains about 16360 unique records. The records contain forecast information by segments, forecast year type (opening, mid-design or design year) and actual count information, if applicable. For analysis purpose, the filters as described in Chapter 3 were applied and we got to 3911 unique records. The data-frame to be analyzed contains project information (unique project ID, type of project, segment ID, roadway functional classification, area type), forecast information (year forecast was produced, forecast year, forecasted and adjusted traffic) and the actual count information (year of observation, count, stations ID).

Based on the nature of the NCHRP 08-110 database, we can select some variables that might dictate future adjustments in the forecasts. These variables are: the type of Project (Improvement Type), the methodology used (Forecast Methodology), roadway type (Functional Class), area type (Area Type Functional Class) and the forecast horizon (difference between year forecast produced and year of opening).

Table 14: Descriptive Variables for Analysis

Variable	Explanation
Forecast Volume	We expect the percent difference from forecasts to be larger for lower volume roads because there is less opportunity for errors to average out.
Functional Class	To test whether accuracy differs for different functional class of roads. The distribution is done on the FHWA defined Functional Classes.
Area Type	To test whether urban or rural areas influence the forecast accuracy.
Type of Project	Distribution of forecast errors across different types of improvement, i.e. resurfacing project, adding lanes, new construction etc. Can be simplified as forecasts on Existing Roads and New Constructions.
Tolls	Relation between toll road forecasts and un-tolled road forecasts.
Opening Year	Projects affected by a recession may have uniformly low forecasts. The Opening Year is taken to be the Year the actual traffic count was taken in our database. The years 2001 and 2008-9 were

	identified as recession years. Judging from the unemployment rate, the years affected by the recession was categorized.
Year forecast produced	To evaluate whether forecast accuracy has improved over the years.
Forecast Horizon	Derived variable from the difference between the Forecast Year and the Year Forecast was Produced. Tests hypothesis that Forecasts are better when the opening year is closer to the year forecast was produced.
Unemployment Rate in Opening Year	To evaluate the effect of recessions on forecast accuracy.
Change in Unemployment Rate	This will be measured as the difference between the unemployment rate in the opening year and the unemployment rate in the year the forecast was produced.
Forecast Methodology	To evaluate the relative accuracy of Trend Based Forecast or Model Based Forecast etc.
Type of Forecaster	To examine differences between forecasts made by DOTs, MPOs, consultants, or others.

In the remainder of this chapter, we examine the overall distribution of percent difference, as well as the percent differences segmented by each of these factors. The codes for cleaning up the data and the analysis itself is available at https://github.com/jawadmhoque/accuracy-assessment repository.

4.3 Overall Distribution

Generally speaking, traffic forecasts have been found to be over-predicting: actual traffic volumes after project has been completed are lower than what has been forecasted, as shown in Figure 5 and Figure 6, which show a right-skewed distribution. The MAPDF is 24.67% at segment level, but this statistic is biased in the sense that multiple segments make up a single project, and a particular error or shortcoming of the method adopted is accumulating over a project. In segment-level, the traffic volumes are off by about 5150 vehicles per day (vpd) on average.

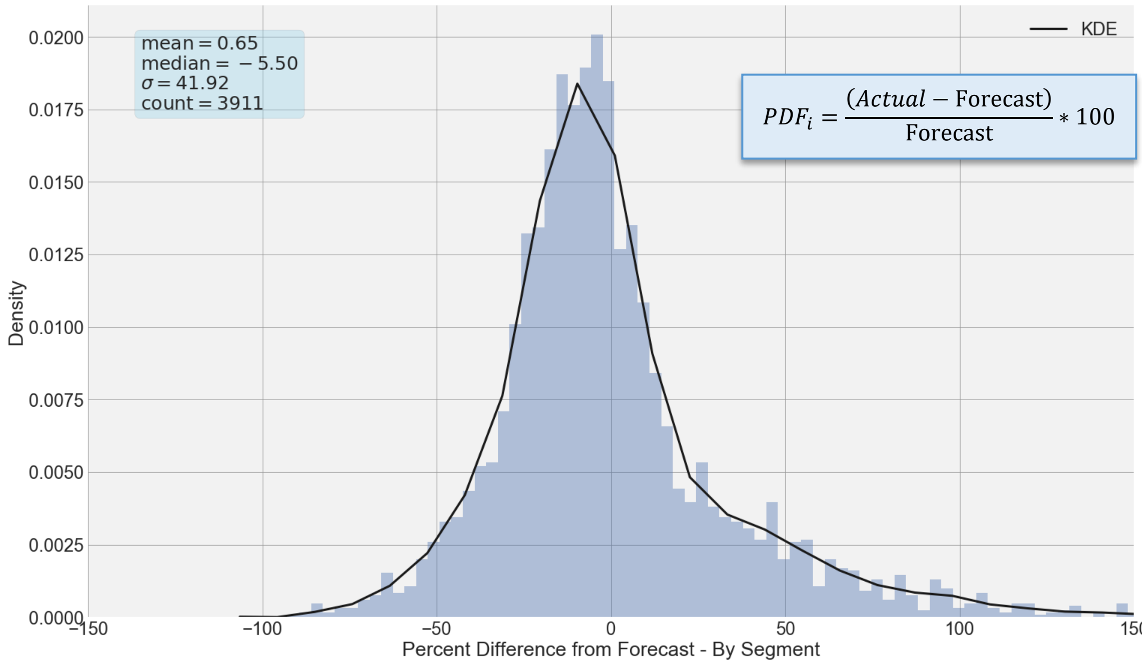


Figure 5: Distribution of Percent Difference from Forecast (Segment Level)

The 3911 unique records/segments are part of 1291 unique projects. Similar to our segment-level analysis, we notice a general over-estimation of traffic across the projects. The distribution of PDF shown in Figure 6 is heavier on the negative side, i.e. actual volumes are generally lower than traffic forecasts. The MAPDF is 17.29% with a standard deviation of 24.81. The Kernel Density Estimator displays an almost normal distribution, albeit with long tails. On an average, the traffic forecasts for a project are off by 3500 vpd.

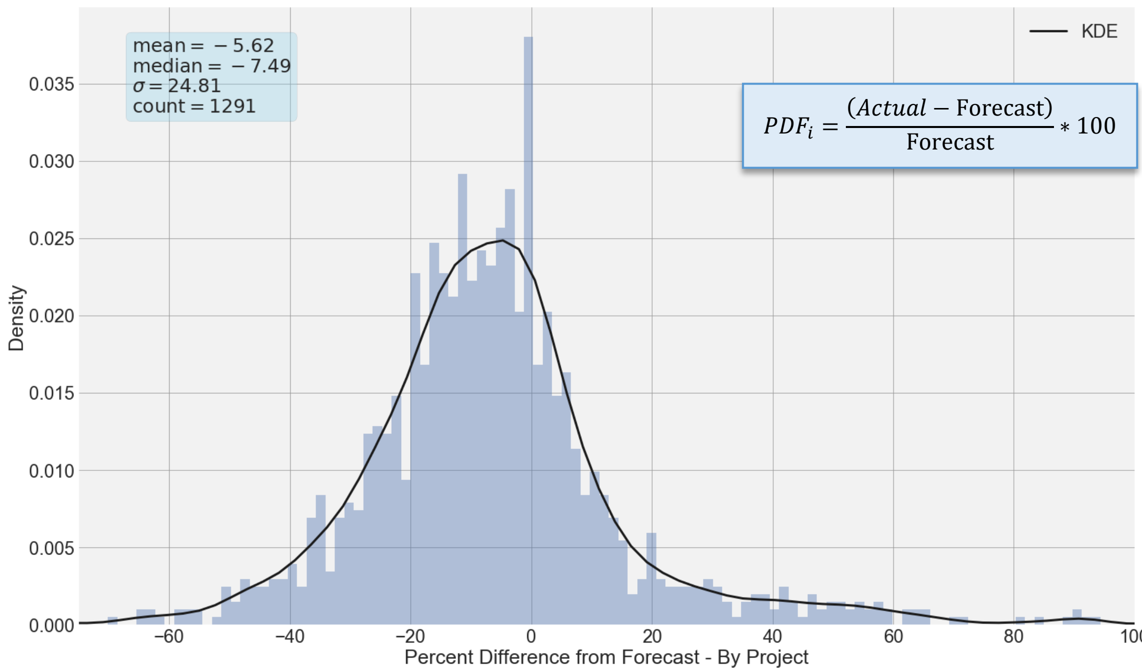


Figure 6: Distribution of Percent Difference from Forecast (Project Level)

We should expect over-predictions because, in many cases, these forecasts are used in design engineering. A design based on over-predicted traffic will be over-built and will not see that extra capacity utilized. On the other hand, if the under-predicted traffic is used as a basis for design, it would mean adding capacity at a later time at a greater cost to meet the demand. This is an example of optimism bias previously noted for toll road traffic forecasts (Bain 2009; Flyvbjerg, Holm, and Buhl 2005).

Table 15: Overall Percent Difference from Forecast

Traffic Forecast Range (AADT)	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Segment Level	3911	24.67	0.65	-5.49	41.92	-44.89	66.34
Project Level	1291	17.29	-5.62	-7.49	24.81	-37.56	36.96

4.4 Forecast Volume

Figure 7 reports the forecast error as a function forecast volume at the segment level. Figure 8 shows it reported at the project level. They are reported separately here because the traffic volume can be quite different for different segments within a project, such as may be the case of a freeway interchange where the mainline freeway volume is much higher than the ramp volumes.

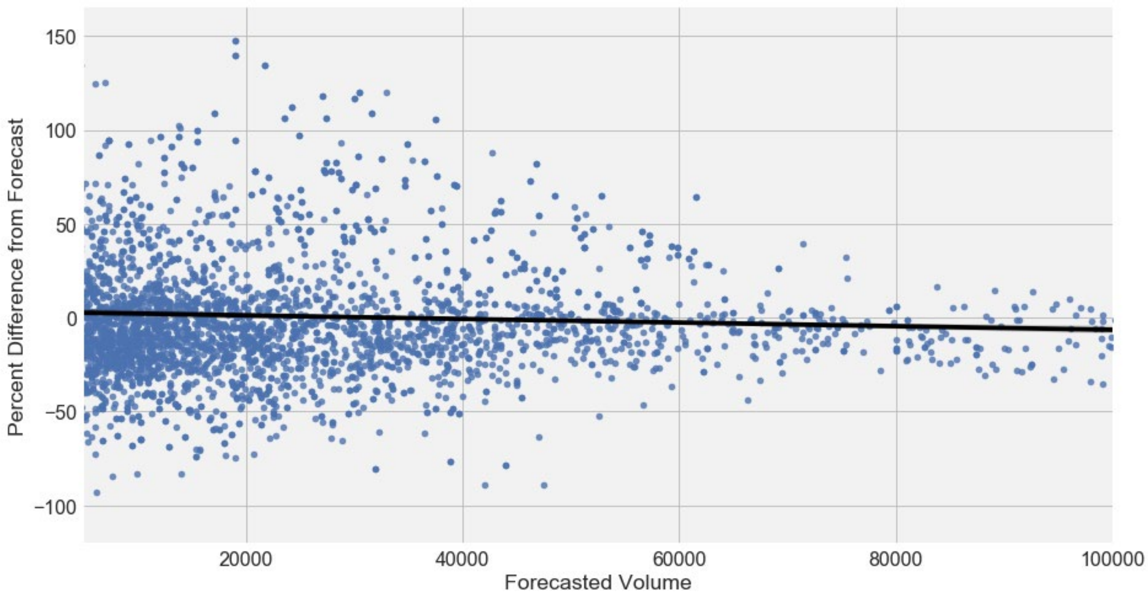


Figure 7: Percent Difference from Forecast as a Function of Forecast Volume (Segment Level)

An interesting observation is the low percentage errors as the traffic volumes increase. This is understandable, since the error percentages were taken as a ratio over the forecasted volume. Unless the actual traffic differs by a large margin, the percentage errors will not have risen to a big amount. The percent difference hover more towards the negative side as we move to the right for higher volume roads. A small number of segments with greater than 80,000 AADT have been under-predicted.

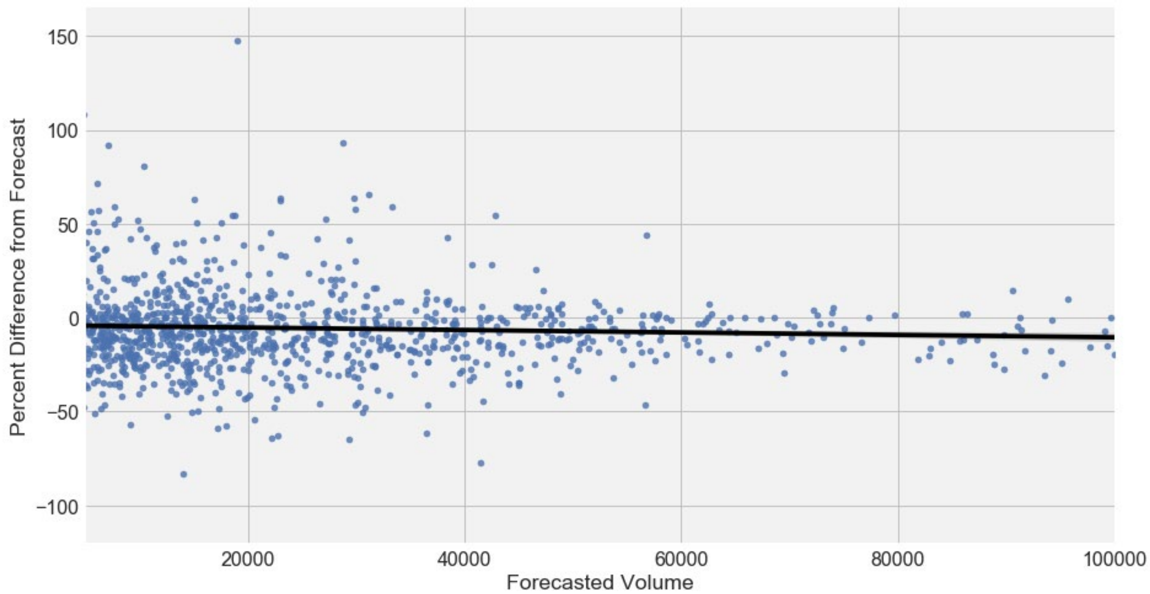


Figure 8: Percent Difference from Forecast as a Function of Forecast Volume (Project Level)

Table 16 and Table 17 show descriptive measures of percent difference of the forecasts by volume group for segments and projects, respectively. The measures represent the spread of the percent difference in forecast, with the Mean, Standard Deviation and 5th and 95th percentile values. The MAPDF value for each category presents how much the actual traffic deviates from the forecast value. Mean is the central tendency of the data. Standard Deviation and the 5th and 95th percentile data represent the spread of the distribution. 90% of the data points fall between the 5th and 95th percentile values.

Table 16: Forecast Inaccuracy by Forecast Volume Group (Segment Level)

Traffic Forecast Range (AADT)	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
(0, 3000]	359	36.17	14.04	-2.22	91.63	-44.78	106.91
(3000, 6000]	419	26.64	3.90	-3.33	38.91	-40.03	83.78
(6000, 9000]	394	24.83	-2.78	-8.93	33.06	-47.90	57.47
(9000, 13000]	465	23.17	-2.54	-6.03	30.11	-44.49	54.98
(13000, 17000]	353	25.31	-0.20	-3.34	34.49	-49.56	76.88
(17000, 22000]	360	25.02	-5.21	-10.40	34.67	-51.54	65.85
(22000, 30000]	415	28.01	3.87	-3.57	37.20	-47.40	77.78
(30000, 40000]	386	25.71	-0.17	-7.92	35.23	-44.64	72.84
(40000, 60000]	410	19.37	2.56	-0.89	26.34	-32.56	53.47
(60000+	350	12.38	-7.14	-6.40	14.98	-28.42	17.50

Table 17: Forecast Inaccuracy by Forecast Volume Group (Project Level)

Traffic Forecast Range (ADT)	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
(0, 3000]	133	24.59	-1.85	-5.75	42.15	-45.01	75.17
(3000, 6000]	142	20.53	-0.37	-4.64	29.74	-36.50	50.33
(6000, 9000]	125	16.75	-5.68	-8.80	21.94	-35.29	36.67
(9000, 13000]	145	15.59	-4.66	-7.29	19.99	-31.34	34.45
(13000, 17000]	143	17.41	-6.20	-6.53	21.61	-37.76	30.65
(17000, 22000]	113	17.98	-5.65	-8.31	25.47	-41.62	37.85
(22000, 30000]	133	19.54	-5.65	-8.47	25.36	-40.31	41.75
(30000, 40000]	115	15.56	-9.78	-10.26	18.23	-39.54	12.26
(40000, 60000]	137	13.18	-8.95	-7.68	16.01	-34.44	7.49
(60000+	105	10.20	-8.96	-7.90	9.90	-24.50	3.68

One observation from Table 17 is that as the forecast volume increases, the distribution of the PDF has smaller spreads in addition to the MAPDF value getting smaller and heavier in the negative side. For example, for forecast volume between 22,000 and 30,000, PDF for 90% of the projects lie between -40.31% and 41.75% with an absolute deviation (MAPDF) of 19.54%. In comparison, 90% of the projects with forecasted traffic between 30,000 and 40,000, have PDF between -39.54% and 12.26% with a MAPDF of 15.56%.

4.5 Functional Class

The distribution of PDF by functional class (Figure 9 and Table 18) are taken at the segment level, since a project may span over roadways of different functional class. Violin plots, as depicted in the figure shows quantitative data with a kernel density estimation of the underlying distribution. The thick black bars represent the 25th and 75th percentile values, in effect depicting the range of values where 50% of the data-points fall in. These reiterate the point made about over-prediction in forecasts: about 75% of the links have negative PDF values for Interstates, Major Arterials and Collectors. About 70% of the Minor Arterial links have been over-predicted.

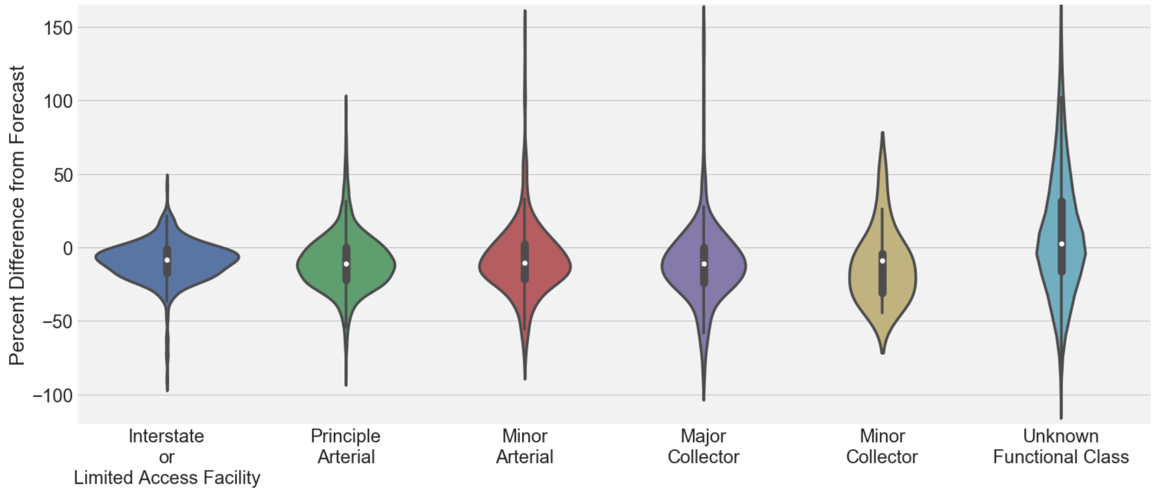


Figure 9: Distribution of Percent Difference from Forecast by Functional Class (Segment Level Analysis)

Compared among themselves, it appears that forecasts for Interstates or Limited Access Facilities fare better than other classes of roadway, both in terms of the absolute deviation and spread (Table 18). 90% of the records of this functional class fall between -27.81% and 10.44%. The spread is a greater for other functional classes (represented by the 5^{th} and 95^{th} percentile values).

Table 18: Forecast Inaccuracy by Functional Class (Segment Level Analysis)

Functional Class	Observations	MAPDF	Mean	Median	Standard Deviation	5^{th} Percentile	95^{th} Percentile
Interstate or Limited Access Facility	434	12.32	-9.21	-8.48	13.58	-27.81	10.44
Principle Arterial	837	16.95	-9.63	-10.89	19.38	-37.51	23.95
Minor Arterial	404	18.92	-8.26	-10.24	24.54	-41.50	29.26
Major Collector	258	20.67	-10.81	-11.10	26.92	-51.11	23.85
Minor Collector	19	22.53	-12.74	-8.66	24.30	-41.43	28.58
Local	1	46.67	46.67	46.67		46.67	46.67
Unknown Functional Class	1958	32.42	10.69	2.68	53.67	-48.75	86.21

4.6 Area Type

The distribution and spread of forecast errors as a function of the area type is presented in Figure 10 and Table 19. Forecasts for both rural and urban areas are mostly over-predicting i.e. actual traffic is less than forecasted (65% of the links in rural area and 72% of links in urban areas).

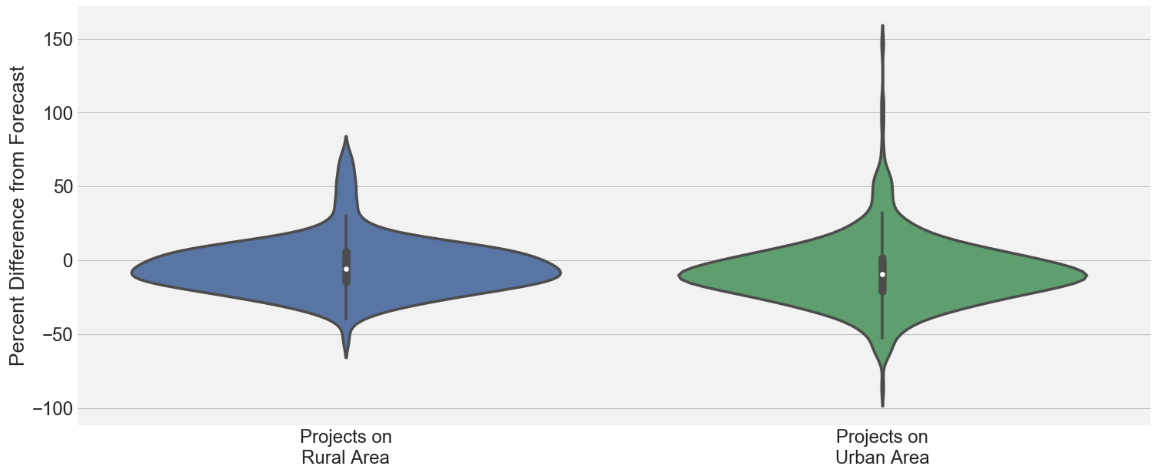


Figure 10: Distribution of Percent Difference from Forecast by Project Area Type (Segment Level Analysis)

The spread for Urban areas (-39.37% to 27.14%) is greater than that for rural areas (-27.93% to 24.72%). The MAPDF values for Rural and Urban areas (14.09% and 17.66% respectively) point to traffic in rural or mostly rural areas have a smaller deviation from predicted.

Table 19: Forecast Inaccuracy by Area Type (Segment Level Analysis)

Area	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Rural or Mostly Rural	210	14.09	-4.02	-5.56	18.22	-27.93	24.72
Urban	543	17.66	-8.05	-9.58	22.32	-39.37	27.14
Unknown Area Type	3047	23.86	-0.12	-5.00	33.89	-47.31	68.05

4.7 Type of Project

As described in Section 3.2.3, the NCHRP 08-110 database has the improvement type of the project as a required field. A lot of the segments/projects don't have any improvement type assigned but we can still unify the types coded in the database in three ways:

1. Improvement on Existing facility: Resurfacing, replacement and adding capacity to existing roadway.
2. New Construction: New general-purpose, dedicated, managed or reversible lane(s) facility and

3. Unknown Project Type.

Among the 1291 projects, our database contains forecast and actual count information on only 28 new construction projects, while projects on existing roadway are 788 in number. About 75% of the projects on existing roadway in the database have error below 0% i.e. over-predicting the traffic. Similar proportions are obtained for New Constructions as well (Figure 11 and Table 20). Compared to aggregated error over all types of project (MAPDF of 17.29%), forecasts for existing roadway have on an average slightly less error (MAPDF of 16.26%). Forecasts for New Constructions are even more accurate with an MAPDF of 10.57%.

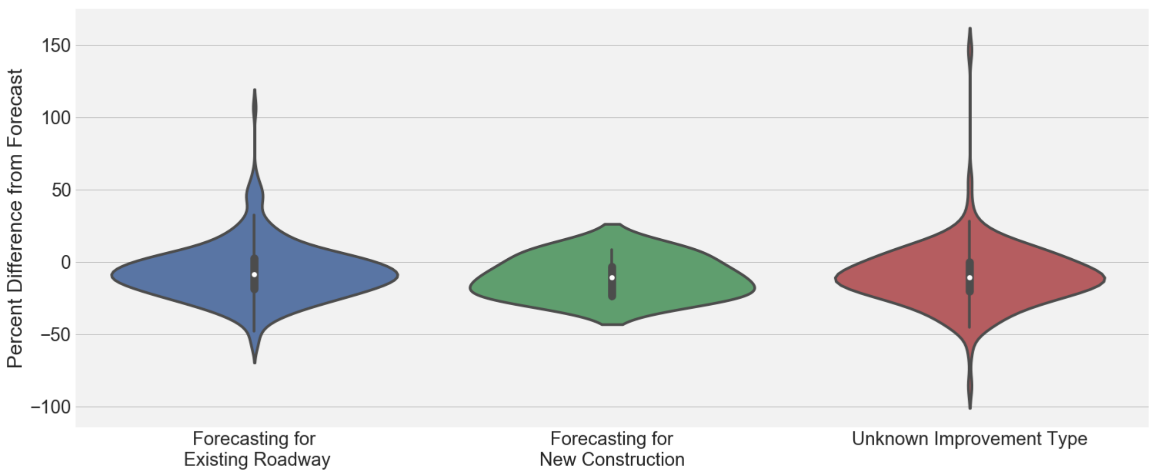


Figure 11: Distribution of Percent Difference from Forecast by Project Type (Project Level Analysis)

The difference in sample sizes make commenting on the relative accuracy of forecasts by project type difficult. But as the percentile values indicate, forecasts for new construction projects have a lower spread than that for existing roadways.

Table 20: Forecast Inaccuracy by Project Type (Project Level)

Project Type	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Existing Road	899	16.26	-5.90	-7.43	23.55	-36.20	29.93
New Facility	28	10.57	-9.22	-8.76	9.54	-19.34	3.83
Unknown Type	364	20.36	-4.64	-7.64	28.38	-43.96	45.95

4.8 Tolls

In our database we didn't have much information about the toll roads. In all, there are forecast information on only 7 roads/links with Static Tolls on 1+ lanes. The MAPDF for the tolled roads is 20.41% with a maximum of 93.38%. The distribution in Figure 12 is not scaled by the number of observation. Table 21 presents the breakdown of the distribution by Toll Type on links.

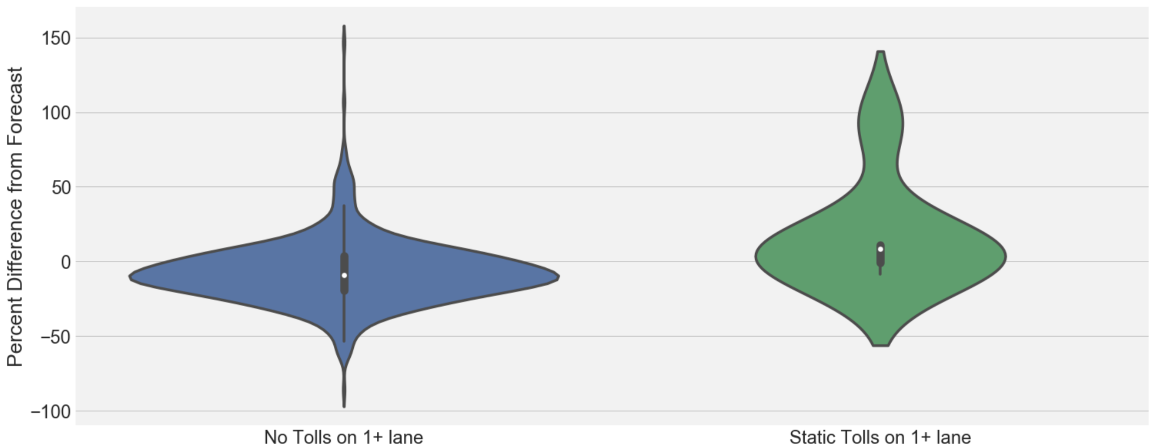


Figure 12: Distribution of Percent Difference from Forecast by Toll Types (Segment-Level Analysis)

Table 21: Forecast Inaccuracy by Toll Type (Segment Level)

Toll Type	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
No Tolls on 1+ lane	3432	23.66	-1.53	-6.55	32.87	-45.9	64.66
Static Tolls on 1+ lane	7	20.41	16.16	8.60	34.96	-7.97	68.85

4.9 Year Forecast Produced

The NCHRP 08-110 database contains projects spanning from 1970s. Forecasts for the projects thus go even before that. In Figure 13 and Table 22 we compare the PDF for forecasts produced in each year. The MAPDF has steadily gone down, in addition to the spread of the distribution getting smaller. Also noticeable is the overall "under-prediction" of traffic for projects that have been forecasted between 1981 to 1990 i.e. actual traffic is more than the forecasted volume. During the next decade (1991-2000), about 55% of the projects for which traffic was forecasted have had more traffic than forecasted. After 2000 however, almost 75% of the projects forecasted have seen less traffic than forecasted with an average absolute deviation of 15.7%.

The improvement over time may suggest that the availability of better data and refinement as well as sophistication of forecasting methodology results in better forecast performance over the years. However, it could be affected by the mix of projects and broader socioeconomic trends. Many of the earlier projects were larger in scale, and the 1970s through 1990s were a time of growing auto ownership, the entry of women into the workforce, and higher VMT per capita. The projects in the 2000s, in contrast, include more routine projects at a time of slower economic growth and slower growth in VMT per capita.

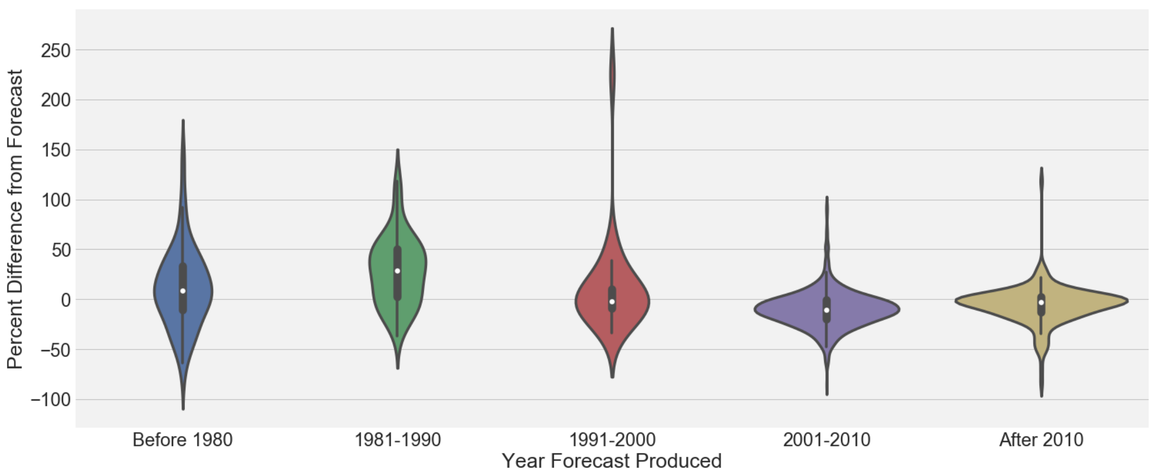


Figure 13: Distribution of Percent Difference from Forecast by the Year Forecast Produced

Table 22: Forecast Inaccuracy by Year Forecast Produced

Year Forecast Produced	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Before 1980	94	30.76	11.25	8.98	39.89	-47.12	83.27
1981-1990	45	34.83	28.21	28.53	34.18	-19.96	86.28
1991-2000	51	23.17	11.13	-1.87	48.07	-24.79	53.56
2001-2010	924	15.79	-9.96	-10.32	18.23	-38.36	15.95
After 2010	177	11.83	-5.36	-2.65	18.81	-38.65	15.62

Analyzing the forecast accuracy for projects on existing roadways, we see similar trends; although after 2010 the MAPDF has gone down from 15.79% in the previous decade to 11.83%. Figure 14 and Table 23 presents the distribution of inaccuracy in projects on existing roads.

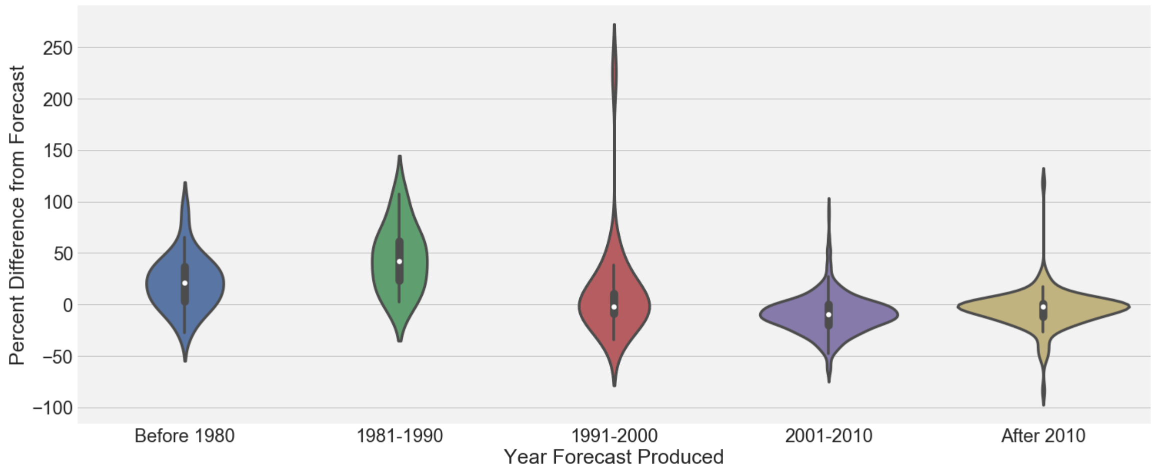


Figure 14: Distribution of Percent Difference from Forecasts for Projects on Existing Roadways by the Year Forecast Produced

Table 23: Forecast Inaccuracy for Projects on Existing Roadways by Year Forecast Produced

Year Forecast Produced	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Before 1980	26	25.59	21.13	21.22	25.87	-14.21	60.72
1981-1990	14	44.76	44.76	42.17	31.76	4.70	96.30
1991-2000	49	23.58	12.12	-1.87	48.74	-23.82	54.21
2001-2010	680	15.78	-9.54	-9.78	18.37	-38.59	18.50
After 2010	130	11.08	-4.51	-1.98	18.68	-32.53	16.39

4.10 Opening Year

The distribution of PDF by the Project Opening Year presented in Figure 15 and Table 24 is a useful indicator of forecast performance over the years. As can be seen, the forecast performance has generally gotten better after 2000, with significantly low MAPDF values than previous decade, as well as smaller spreads. Most of the projects (about 78%) that have opened to traffic between 1991 to 2002 have had more traffic than forecasted. Percent Difference from 2003 to 2008 are more evenly spread (90% data points between -36.82% and 33.46%) while after 2012, actual count has been generally less than the forecasted value (78% of the projects).

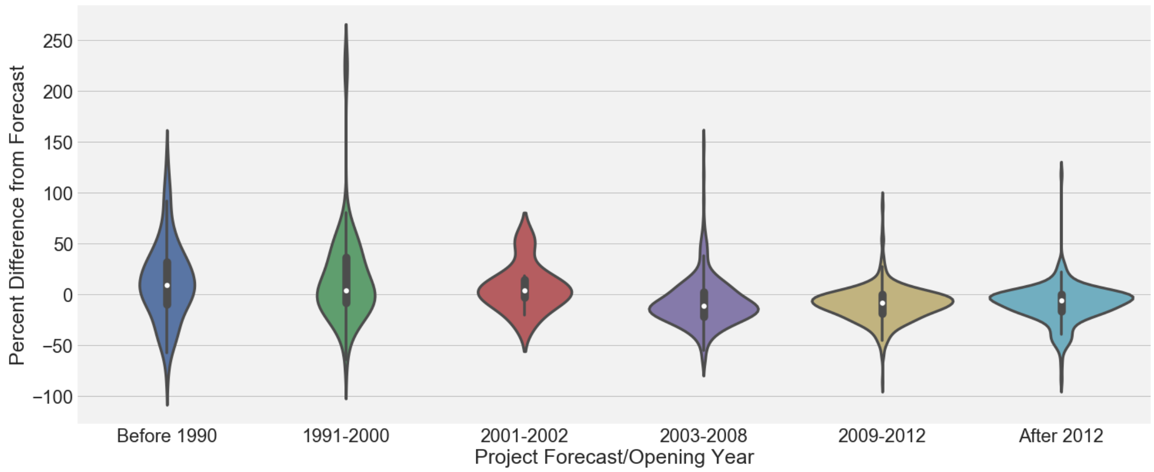


Figure 15: Distribution of Percent Difference from Forecast by Opening Year of Project

The opening years have been categorized to assess the effect of recession (recession in 2001 and the great recession on 2008-09) on forecast performance. It is assumed that the 2001 recession would affect unemployment rate till 2002 and the great recession till 2012, based on the unemployment rate for the years. One thing to notice here is that during and after the recession years, the actual traffic has been lower than usual. The median values (corresponding to 50th percentile value) give a good approximation, as 50% of the projects opened since 2012 have traffic at least 5.78% less than the forecasted value.

Table 24: Forecast Inaccuracy by Project Opening Year

Opening Year	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Before 1990	92	30.14	12.98	9.64	38.24	-43.71	89.49
1991-2000	72	28.09	15.83	3.74	45.17	-28.66	62.88
2001-2002	15	15.65	6.69	3.74	22.50	-22.86	51.82
2003-2008	351	18.92	-7.98	-11.52	23.76	-36.82	33.46
2009-2012	512	14.22	-9.21	-8.46	17.08	-35.07	12.25
After 2012	249	13.56	-8.73	-5.78	18.41	-42.71	13.45

Again, it is not clear the degree to which the differences observed here are a function of different forecasting methods, events in the real world, or a mix of the two.

Looking strictly at the projects done on existing roadways, a similar distribution is observed. The ranges have become tighter, with a lower MAPDF value (except for the projects opening between 1991 and 2000). The distribution and statistical results are given in Figure 16 and Table 25.

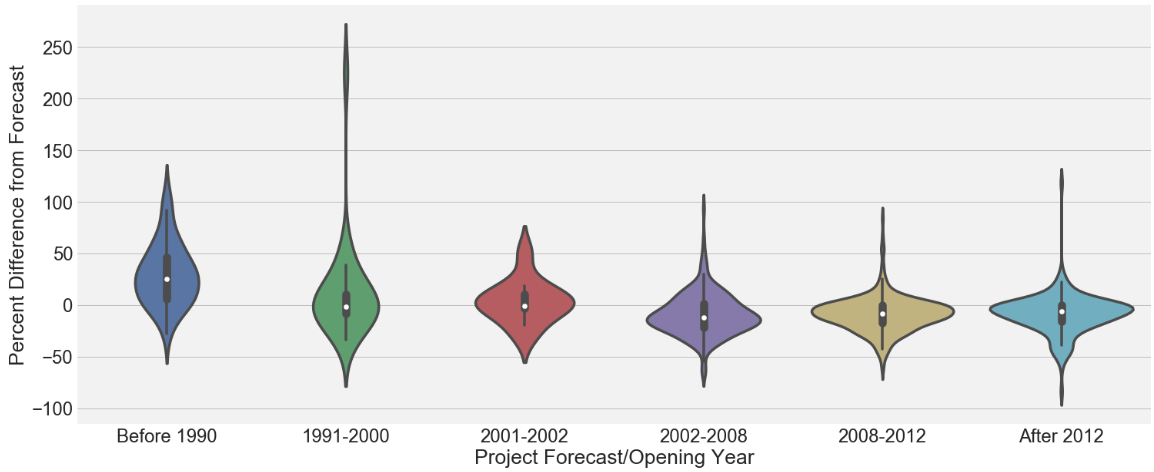


Figure 16: Distribution of Percent Difference for Projects on Existing Roadways by Opening Year of Project

Table 25: Forecast Inaccuracy for Projects on Existing Roadways by Opening Year

Opening Year	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Before 1990	40	32.30	29.40	25.32	29.93	-11.69	90.59
1991-2000	49	23.58	12.12	-1.87	48.74	-23.82	54.21
2001-2002	11	13.88	3.47	-0.75	20.88	-24.81	34.60
2003-2008	247	17.69	-9.21	-11.94	20.24	-35.99	20.27
2009-2012	373	13.95	-8.82	-8.44	16.82	-35.23	13.72
After 2012	179	13.68	-8.65	-5.78	19.08	-42.45	14.12

4.11 Forecast Horizon

Another question that comes to mind while evaluating the accuracy is whether the number of years elapsed between the time forecast was produced to the year project was opened has a bearing on the accuracy. As evident from Figure 17 and Table 26, the average of the absolute PDFs increases as the number of years elapsed increases, except for the same-year projections. The difference in years introduces other variables like micro and macro economy, change in land use and fuel price etc. that can directly affect the traffic. These are all variables that are difficult to predict, and their effect is evident. This finding is consistent with findings by Bain (2009) who identified the critical dependency of longer-term forecasts on macro-economic projections. According to Standard and Poor's Studies (2002-2005)- "*A number of comments were recorded about the relationship between economic growth and traffic growth; concerns being raised about* *traffic forecasts—particularly over longer horizons—relying on strong and sustained economic growth assumptions that resembled policy targets rather than unbiased assessments of future economic performance*."

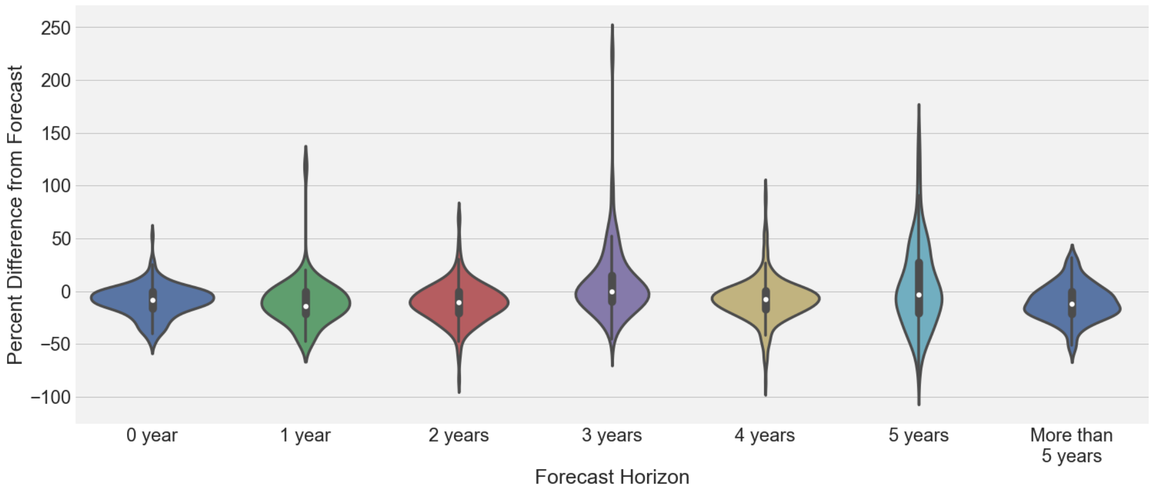


Figure 17: Distribution of Percentage Difference from Forecast by Forecast Horizon

Forecasts that go beyond 5 years in the future tend to have a wider spread and higher PDF (90% of the data point fall within -44.73% to 72.07% with a MAPDF of 29.55%).

Table 26: Forecast Inaccuracy by Forecast Horizon

Forecast Horizon (Years)	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
0	165	20.10	8.08	0.00	34.77	-25.18	57.71
1	206	12.88	-9.20	-8.12	14.64	-36.32	11.38
2	340	15.23	-7.79	-7.64	19.93	-40.26	20.38
3	251	16.25	-10.36	-10.74	18.49	-37.02	17.29
4	131	16.05	-10.36	-12.16	16.87	-35.43	20.19
5	67	16.82	-10.44	-13.82	22.23	-43.99	13.40
5+	131	29.55	4.71	-3.13	39.47	-44.73	72.07

A point on concern in this analysis must be why the MAPDF value as well as the range of forecast error is higher for a forecast horizon of 0 year. 50% of the observation fall on either side of 0% error.

4.12 Unemployment Rate in Opening Year

The Unemployment Rate data was pulled from the Bureau of Labor Statistics at the State level, and then matched with the year the actual traffic count was taken. For European projects it is measured at the national level. The rates were categorized into 7 classes or ranges and the distribution of PDF is presented in Figure 18. Except for Unemployment Rate below 3, PDF hovers in the negative side i.e. over-prediction for all other ranges. For unemployment rate between 1 to 3, the actual traffic is more than the forecasted volume for most of the case, but this statistic should be taken with a grain of salt because of the small sample size.

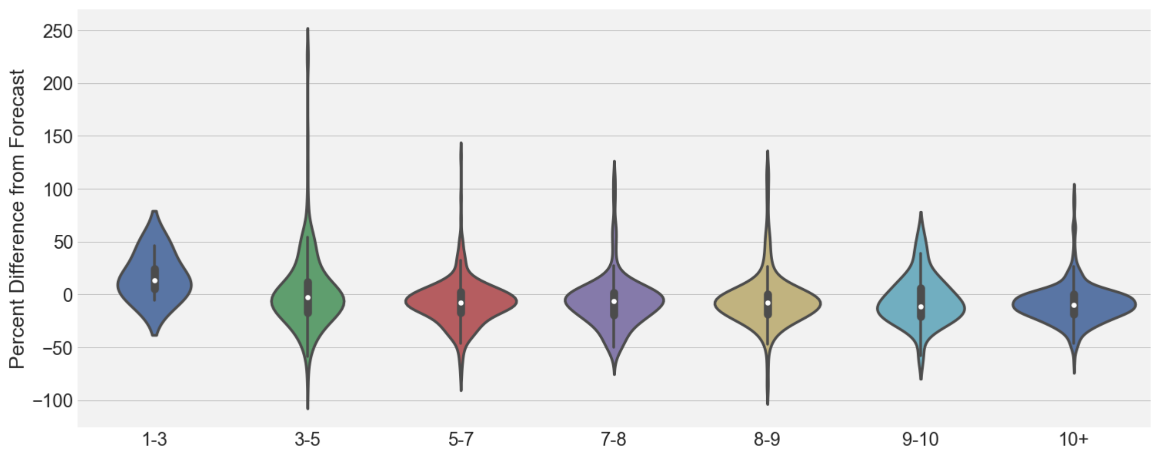


Figure 18: Distribution of Percent Difference from Forecast by Unemployment Rate in Opening Year

Comparing between the ranges, unemployment rate between 4 and 5 seems to produce the maximum absolute deviate from forecast volume. Other ranges hover close to the overall average. Breakdown of the statistics is given in Table 27.

Table 27: Forecast Inaccuracy by Unemployment Rate in the Opening Year

Unemployment Rate	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Up to 3	4	19.44	16.73	13.08	21.70	-3.21	41.78
3-5	229	22.95	2.13	-2.84	36.05	-40.20	55.83
5-7	371	16.10	-7.35	-7.68	21.30	-39.70	26.86
7-8	128	17.30	-7.05	-6.45	24.00	-43.19	26.12
8-9	168	17.07	-5.41	-7.51	24.68	-33.34	35.09
9-10	35	18.17	-5.15	-11.22	22.33	-28.14	39.05
10+	356	14.90	-8.68	-9.64	18.08	-34.43	19.60

4.13 Change in Unemployment Rate

To assess the impact of change in unemployment rate on forecast inaccuracy, we took the difference of Unemployment Rate between the Project Opening Year and Year Forecast was Produced. At least 70% of the project for which the unemployment rate changed by at least ±4% exhibited actual traffic less than the forecast value. The distribution of PDF is presented in Figure 19 and Table 28.

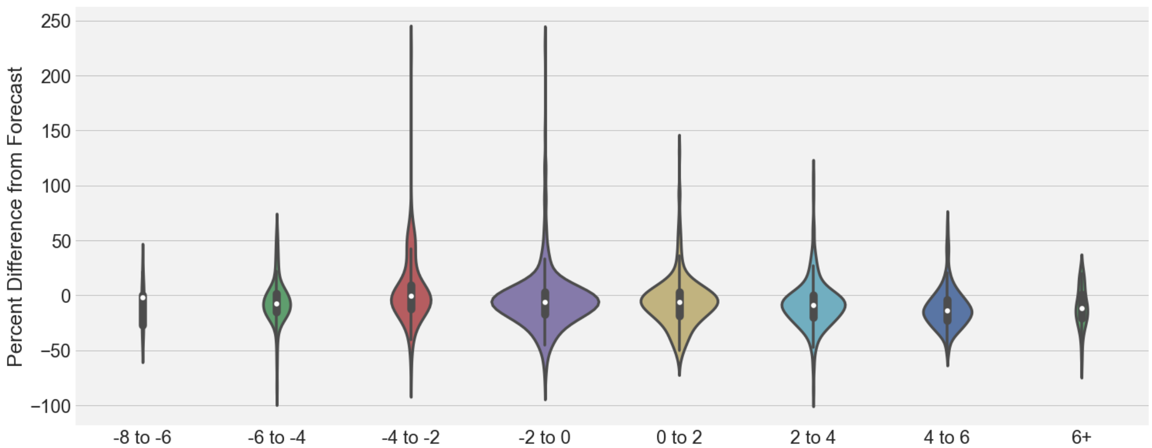


Figure 19: Distribution of Percent Difference by Change in Unemployment Rate from Forecast Year and Opening Year

An interesting, but not quite unexpected, observation is the spread of the distribution for cases where the Unemployment Rate increased in the opening year from the year forecast was produced by at least 2 points. 90% of the projects fall between -36.1% to 26.67% for change of 2-4% and -35.26% to 18.78% for change of 4-6%. With the increase of unemployment rate, it stands to reason that the actual traffic would be less. The possibility of under-prediction would thus get even lower, resulting a narrower range on the positive side.

Table 28: Forecast Inaccuracy by Change in Unemployment rate

Change in Unemployment Rate	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
(-8, -6]	8	15.01	-8.69	-2.02	19.29	-32.69	15.29
(-6, -4]	93	14.91	-5.63	-7.18	20.30	-31.30	31.45
(-4, -2]	136	19.21	4.45	-0.67	31.39	-30.61	54.60
(-2, 0]	367	17.64	-4.27	-6.16	27.88	-38.82	36.58
(0, 2]	263	16.8	-6.00	-6.32	23.27	-40.58	30.62
(2, 4]	217	17.05	-8.01	-8.63	22.12	-36.09	26.67
(4, 6]	166	17.54	-11.75	-13.94	17.80	-35.26	18.78
6+	41	17.1	-10.51	-11.52	17.96	-36.00	19.50

4.14 Forecast Method

One derivative of the Large-N analysis is assessing the performance of the tools at disposal for the state DOTs and MPOs. For project level traffic forecasting, NCHRP Report 765 examines different methods that are in use and presents a guideline for employing those. But one question should arise: does the forecast performance depend on the method used? As a follow up question, is a certain type of forecast methodology better for a certain type of project? Or even a certain type of roadway?

In the NCHRP 08-110 database, a field is specified to record the method used to forecast the traffic for a project. The coded methodologies were: Traffic Count Trend, Population Growth Rate, Regional Travel Demand Model, Project-Specific Travel Demand Model, Professional Judgment and Unknown Methodology. Professional Judgement refers to the usage of a combination of count trend and volume from demand model, as the forecaster saw fit. We have run into the problem of missing data here again, as 676 of the projects in our database have no data regarding the method used to forecast the traffic. Distribution of inaccuracy is presented in Figure 20 and Table 29.

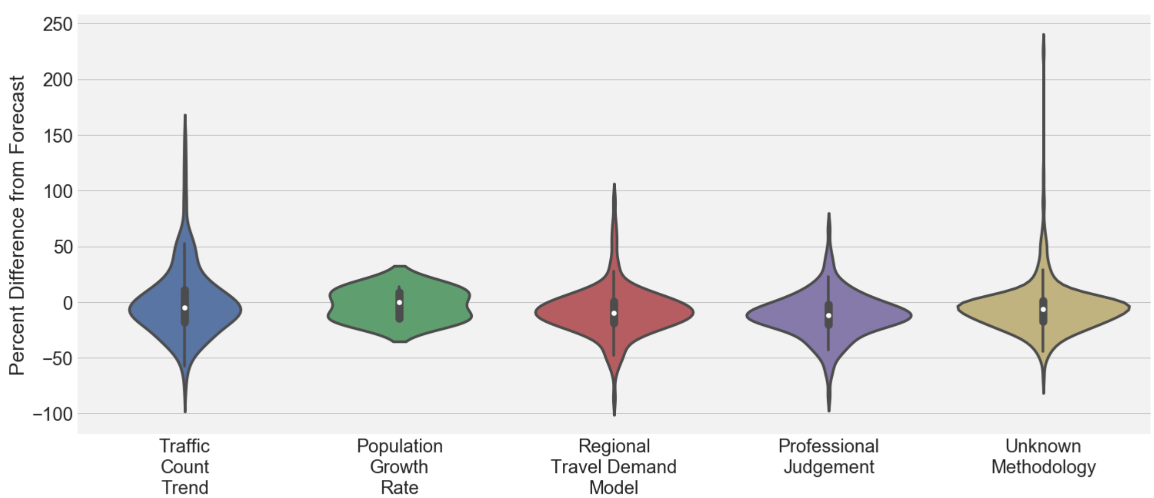


Figure 20: Distribution of Percent Difference by Forecast Methodology

Table 29: Forecast Inaccuracy by Forecast Methodology

Forecast Methodology	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
Traffic Count Trend	252	22.21	-0.10	-5.22	31.24	-39.34	55.06
Population Growth Rate	7	11.32	-2.18	-0.35	13.56	-16.43	13.89
Regional Travel Demand Model	179	16.88	-8.42	-9.75	21.76	-44.91	27.16
Professional Judgement	177	17.84	-11.77	-11.94	19.87	-43.11	18.52
Unknown Methodology	676	15.49	-5.36	-6.45	23.67	-34.39	29.49

Looking at a glance to the distribution of error by forecast methodology (Table 29) we can say that forecasts done by Travel Demand Models are more accurate comparing the MAPDF values (MAPDF of Travel Demand Model is 16.88%, compared to 22.21% of Traffic Count Trend). But it does not accurately portray the picture. As we know, trend analysis cannot be used on all types of projects while models can be used on virtually any type of project.

4.15 Type of Forecaster

The distribution of forecast inaccuracy by the forecaster is presented in Figure 21 and Table 30. As can be seen, 90% of the projects forecasted by State DOTs fall in the range of -44.94% and 54.32%. 55% of these projects are over-predicted. The spread for forecasts done by Consultants is lower (90% of the projects lie between -35.83% and 31.42%), as well as the mean absolute deviation (MAPDF of 17.36% compared to 21.47% for State DOT produced forecasts)

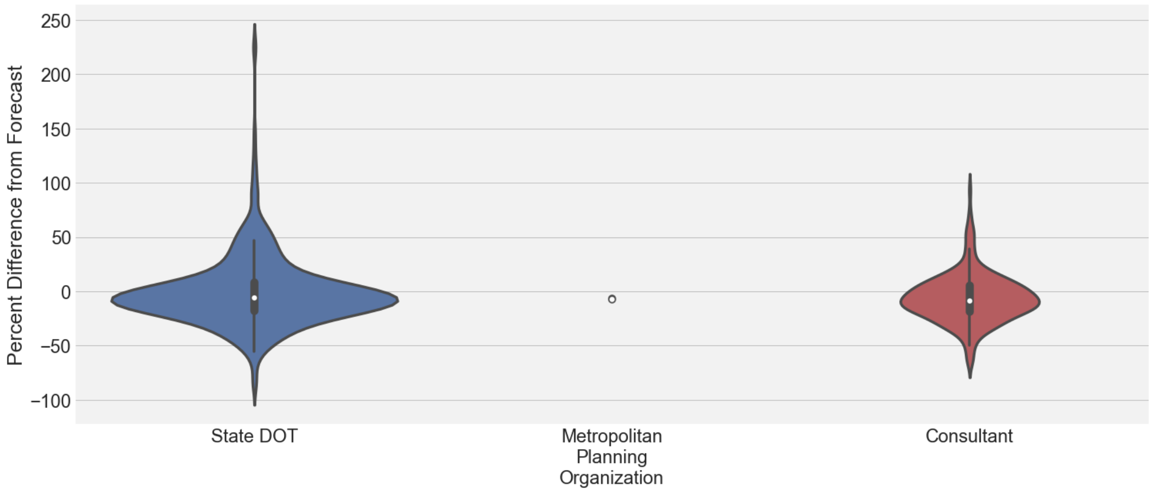


Figure 21: Distribution of Percent Difference by Type of Forecaster

Table 30: Forecast Error by Type of Forecaster

Forecasting Agency	Observations	MAPDF	Mean	Median	Standard Deviation	5th Percentile	95th Percentile
State DOT	489	21.47	-0.89	-5.58	32.34	-44.94	54.32
Metropolitan Planning Organization	2	6.86	-6.86	-6.86	0.90	-7.43	-6.29
Consultant	237	17.36	-6.36	-8.20	22.13	-35.85	31.42

4.16 Effect on Number of Lanes

There is an old axiom that traffic forecast only need to be accurate to within half a lane. To test the extent to which we meet this standard, we calculated the Number of Lanes required for forecasted traffic and the actual traffic, assuming the same Level of Service.

Comparing the two numbers, we found 36 links out of the 3911 (1.0%) that required an additional lane to allow the traffic to flow at the forecasted LOS. This such small number reinforces our interpretation of over-prediction in traffic forecast. As for these 36 links, if the assumptions regarding the number of lanes hold true, the LOS would get worse. 5 of the 36 are Minor Arterials, the rest are Interstate and Major Arterials (16 each).

Conversely, analyzing for the links that over-estimate the traffic by an amount such that they could do with a lesser number of lanes per direction, we get to 158 links (4.2%). 92 of such links are Interstate, 64 are Principle Arterials and the rest are Minor Arterials.

CHAPTER 5. CONCLUSION

5.1 Summary of Findings

The unknown accuracy of un-tolled roads in the US was the focal point of this study. Analyzing the database prepared as part of the NCHRP funded project, this study identifies several variables that can affect the forecast performance. Revisiting the original research questions, we can offer the following conclusions:

What is the distribution of forecast errors across the sample as a whole?

The forecast errors are best summarized by the distribution shown in Figure 22.

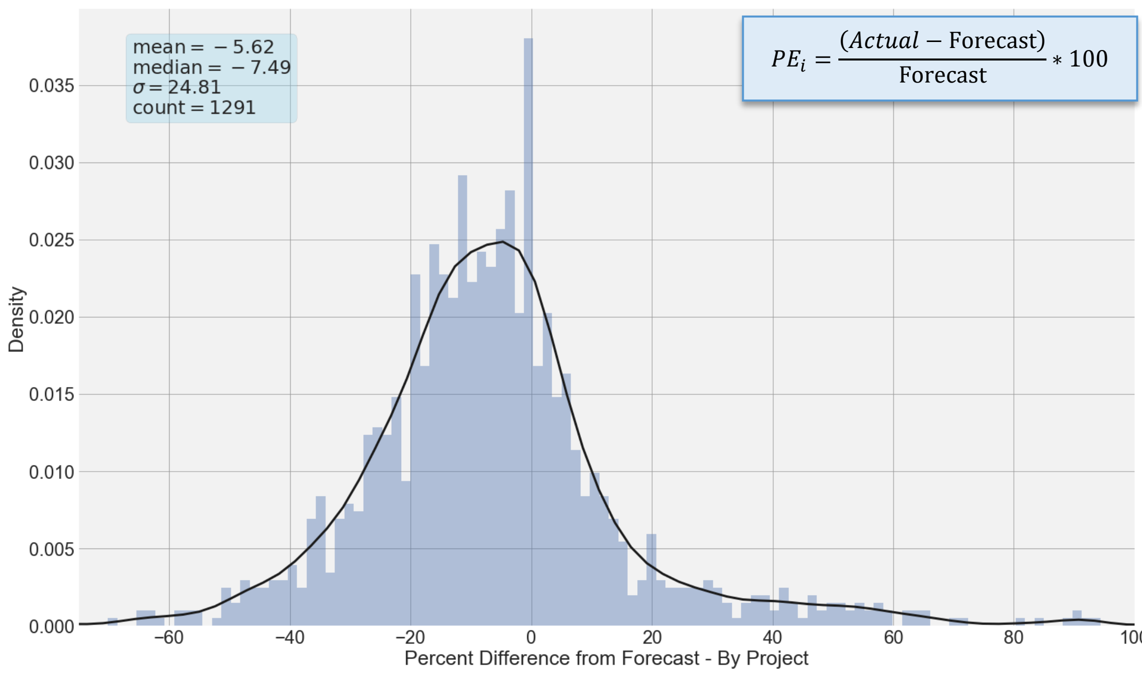


Figure 22: Distribution of Percent Difference from Forecast (Project Level)

Forecast Errors show a significant spread, with a mean absolute percent difference of 25% at the segment level and 17% at a project level. 90% of segment forecasts fall within the range -45% to +66%, and 90% of project level forecasts fall within the range of -37% to +37%.

Can we detect bias in the forecasts?

Yes. Actual ADT is about 6% lower than forecast ADT, and this difference is statistically significant. The fact that most of the projects have less traffic than forecast points to the existence of optimism bias. It is a matter of debate how much of it is actually intentional since most of the forecasts in the database are used for design engineering.

Can we enumerate the sources of forecast error as hypothesized in previous researches?

Several factors are found to affect this bias, including economic conditions, forecast horizon, and facility type. Traffic forecasts show a modest bias, with actual ADT about 6% lower than forecast ADT. The precise number depends upon which metric is used, but the results are in a similar range. The mean percent difference from forecast is +0.65% at a segment level and -5.65% at a project level. The median percent difference from forecast is -5.5% at a segment level and -7.5% at a project level. The difference between the mean and median values occurs because the distribution is asymmetric; actual values are more likely to be lower than forecast, but there is a long right-hand tail of the distribution where a small number of projects have actual traffic much higher than forecast.

We found traffic forecasts to be more accurate for higher volume roads. For example, for segments with 60,000 ADT or more, the MAPDF is 12.4% compared to 24.7% overall.

Traffic forecasts are also more accurate for higher functional classes, over and above the volume effect described above. The actual volumes on lower-class roads are more likely to be lower than the forecasts. These challenges may be due to limitations of zone size and network detail, as well as less opportunity for inaccuracies to average themselves out on larger facilities.

Traffic forecasts become less accurate as the forecast horizon increases, but the result is asymmetric, with actual ADT more likely to be higher than forecast as the forecast horizon increases.

Regarding the performance of forecasting techniques, we found regional travel models producing more accurate forecasts than traffic count trends. The MAPDF for regional travel models is 16.9% compared to 22.2% for traffic count trends.

Traffic forecasts have improved over time. This can be observed both in our assessment of the year the forecast was produced and in the opening year. Forecasts for projects that opened in the 1990s were especially poor, exhibiting mean volumes 35% higher than forecast, with a MAPDF of 32%.

We find that 95% of forecasts reviewed are “accurate to within half of a lane”. We find that for 1% of cases, the actual traffic is higher than forecast and additional lanes would be needed to maintain the forecast level-of-service. Conversely, for 4% of cases, actual traffic is lower than forecast, and the same Level of Service could be maintained with fewer lanes.

5.2 Limitations

It is important that we note a limitation of this study: the data used here are not necessarily a random or representative sample of all traffic forecasts. They were assembled based on availability and shared from different agencies and past researchers examining the topic. As a result, the data contain missing fields that are different, depending on the source. Additionally, it analyzes data from only 6 states. An even more representative result can be obtained if data from states that are experiencing rapid economic growth, particularly the mid-west region of the US were included in the analysis.

We know that the data provided by different agencies comes from different time periods, with different mixes of projects. From what we know examining the data, routine projects such as repaving and minor improvements are more likely to be recorded in more recent years, as records of those projects are less likely to be maintained over a span of decades. While we might think that forecasts get better over time because we now have access to better data, more computational power and better models, it may also be that the forecasting task has become easier over time. Infrastructure budgets are constrained, and states today build fewer big projects. The span between the 1970s and the 1990s was one of growing auto ownership and an increasing share of women in the workforce, which logically would lead to more VMT per capita and measured volumes higher than forecast, whereas both trends had largely played out by the 2000s. It is difficult to disentangle these factors, and we are left to speculate: if we are interested in drawing an uncertainty window around our present-day forecasts, how much credit should we take for recent improvements in forecasts?

5.3 Future Research

This study bridges the gap of unknown traffic 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 traffic forecasts.

A natural continuation of the work presented here is analyzing how the forecast performance can change with the increase and decrease of the variables. A form of it has been explored, by means of quantile regression. But this is beyond the scope of this thesis.

A limitation in the analysis is the lack of statistical tests to determine the effects of the categorical variables and test how different these variables affect the PDF. These tests are called experiments. The analysis of the data from a statistically designed experiment provides answers to the hypothesis in the experimental study. For example, we will be able to test:

1. All “treatment”, or any specific experimental condition applied to the response variable i.e. the categorical variables have the same effect,
2. A particular variable in a class of variables affects the response variable more than others and
3. How large are certain variables, or a group of variables than others?

The hypothesis for testing the equality of the variables can be tested by the Analysis of Variance (ANOVA test). The effect of each individual variables can be tested by different tests for Contrasts in the Generalized Linear Model (GLM) procedure.

The sources of forecast errors identified in this research leads to the question of quantifying the effect of getting them correct for a project. This would entail obtaining of the model used to forecast traffic for a project and changing the variables to reflect the actual or correct situation. For example, this study shows that unemployment rate in the opening year may be a factor for error. We take a particular project, document the assumptions and re-run the model used with an updated value of unemployment rate (provided it is in the model, of course) to record the change in accuracy level. Such analysis would pave way to specify which of the exogenous variables need most attention during forecast.

As with previous research, this study suffers from data limitations. 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.

Another interesting area that can be explored is getting the most accurate forecast with the least amount of data. We have identified the sources of error in this study and we can analyze the effect of these factors. Taking these together, we can possibly ascertain the minimum amount of information necessary to get to the most accurate forecast, or as much accuracy as we want. Then again, this requires an answer to the question: how accurate do we want our forecasts to be?

REFERENCES

Anderson, Michael, Walter C. Vodrazka, and Reginald R. Souleyrette. 1998. "Iowa Travel Model Performance, 20 Years Later." In , 585–95. ASCE

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

Bain, Robert, and Jan Willem Plantagie. 2004. "Traffic Forecasting Risk: Study Update 2004." http://www.people.hbs.edu/besty/projfinportal/s&p_traffic_risk_2004.pdf.

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.

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

CDM Smith, Alan Horowitz, Tom Creasy, Ram M Pendyala, Mei Chen, National Research Council (U.S.), Transportation Research Board, et al. 2014. *Analytical Travel Forecasting Approaches for Project-Level Planning and Design*. Washington, D.C.: Transportation Research Board.

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.

FDOT, Mohammad. 2014. "Project Traffic Forecasting Handbook."

Federal Transit Administration. 2016. "Guidance on Before-and-After Studies of New Starts Projects." Text. FTA. April 4, 2016. https://www.transit.dot.gov/funding/grant-programs/capital-investments/guidance-and-after-studies-new-starts-projects.

Federal Transit Administration, and Vanasse Hangen Brustlin. 2008. "The Predicted and Actual Impacts of New Starts Projects -- 2007: Capital Cost and Ridership."

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, B., 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. 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 – 597. https://doi.org/10.1068/b32111.

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.

Highways England. 2015. “Post Opening Project Evaluation (POPE) of Major Schemes: Main Report.”

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.

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.

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.

MacKinder, I. H., and S. E. Evans. 1981. “The Predictive Accuracy of British Transport Studies in Ubran Areas.” *Transport and Road Research Laboratory*. https://trid.trb.org/view.aspx?id=179881.

Marlin Engineering. 2015. “Traffic Forecasting Sensitivity Analysis.” TWO#13.

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.

National Oceanic and Atmospheric Administration. 2010. “Hurricane Forecast Improvement Program: Five-Year Strategic Plan.” http://www.hfip.org/documents/hfip_strategic_plan_yrs1-5_2010.pdf.

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. 2014a. “Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review.” *Transport Reviews* 34 (4): 540–57. https://doi.org/10.1080/01441647.2014.926428.

Nicolaisen, Morten Skou, and Petter Næss. 2015. “Roads to Nowhere: The Accuracy of Travel Demand Forecasts for Do-Nothing Alternatives.” *Transport Policy* 37 (0). https://trid.trb.org/view/2015/C/1334458.

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.

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.

Pickrell, D. H. 1989. “Urban Rail Transit Projects: Forecast Versus Actual Ridership And Costs. Final Report.” https://trid.trb.org/view.aspx?id=299240.

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

Toepfer, Fred. 2015. “Presentation to TRB Travel Demand Forecasting Committee Meeting.” January 13.

Tsai, Chi-Hong Patrick, Corinne Mulley, and Geoffrey Clifton. 2014. “Forecasting Public Transport Demand for the Sydney Greater Metropolitan Area: A Comparison of Univariate and Multivariate Methods.” *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice* 23 (1): 51.

U.S. Department of Transportation: Federal Transit Administration. 2003. “Predicted and Actual Impacts of New Starts Projects: Capital Cost, Operating Cost and Ridership Data.”

Vassallo, J.M, and M. Baeza. 2007. “Why Traffic Forecasts in PPP Contracts Are Often Overestimated?” *EIB University Research Sponsorship Programme*.

Voulgaris, Carole Turley. 2017. “Crystal Balls and Black Boxes: Optimism Bias in Ridership and Cost Forecasts for New Starts Rapid Transit Projects.” University of California Los Angeles. https://escholarship.org/content/qt1478s66b/qt1478s66b.pdf.

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

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.

VITA

Jawad Mahmud Hoque

Education:

August 2017- Current	University of Kentucky M.Sc. in Civil Engineering
April 2016	Bangladesh University of Engineering and Technology, Dhaka, Bangladesh B.Sc. in Civil Engineering

Professional Positions:

August 2017-Current	Graduate Research Assistant Department of Civil Engineering, University of Kentucky
August 2018-Current	Graduate Teaching Assistant Department of Civil Engineering, University of Kentucky
September 2016-June 2017	Project Associate Transportation Unit, The World Bank

Awards and Certification:

- Part of the 3-member Champion team to the Regional Traffic Bowl Competition, Southern District Institute of Transportation Engineers, held on Arlington, Virginia on April, 2019.
- Certified Associate in Project Management (CAPM), CAPM license no. 197322.

Publications:

- Tanzen, R., Hoque, J. M., & Islam, M. S. (2018) Predicting Highway Crash Frequency: A Review of Statistical Regression Models. *Journal of Modern Science and Technology*, 6 (1), pp 87-99.