UNIVERSITY OF CALIFORNIA

Los Angeles

Crystal Balls and Black Boxes:

Optimism Bias in Ridership and Cost Forecasts

for New Starts Rapid Transit Projects

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Urban Planning

by

Carole Turley Voulgaris

2017

ABSTRACT OF THE DISSERTATION

Crystal Balls and Black Boxes:

Optimism Bias in Ridership and Cost Forecasts

for New Starts Rapid Transit Projects

by

Carole Turley Voulgaris

Doctor of Philosophy in Urban Planning

University of California, Los Angeles, 2017

Professor Brian D. Taylor, Chair

Several studies have observed an optimistic bias in cost and ridership forecasts for rail transit projects around the globe, which has led to billions of dollars of public investment in projects that have not performed as promised. This bias has been a major cause of concern for project stakeholders, including the Federal Transit Administration (FTA), which has spent an average of over $3 billion each year over the past two decades on new rail transit projects in the United States through its Capital Investment Grants program, commonly known as New Starts. Partly in response to credibility concerns raised by forecast bias, the FTA has made changes to the New Starts program over the years. However, there has been no research to date that has examined how these changes in the New Starts program have influenced forecast accuracy for rail transit projects that receive funding.

This study addresses that gap in the literature through a mixed-methods approach involving semi-structured interviews with thirteen transit planning and forecasting professionals and a quantitative analysis of 67 completed transit projects to determine whether and to what extent forecast accuracy has changed over time and what changes in federal policy and transit planning practice might explain these changes.

I find that there have been steady improvements over time in the accuracy of ridership forecasts and cost estimates for New Starts projects. The improvement in ridership forecast accuracy can be explained in part by shorter project construction durations and a shift over time in the perceived purpose of forecasting from (1) project promotion to (2) fairness of competition to (3) use in local decision-making. Some of the improvement in cost estimate accuracy can be explained in by changes in project characteristics, particularly a tendency towards more modest projects representing incremental changes to the transit network.

This analysis of forecast bias in transit planning gives us reasons for optimism regarding the future of optimism bias in cost and ridership forecast accuracy, since forecasts appear to be on a long-term trajectory toward more accuracy and less bias.

The dissertation of Carole Turley Voulgaris is approved.

Joan Walker

Martin Wachs

Brian D. Taylor, Committee Chair

University of California, Los Angeles

2017

DEDICATION

To my family: Phil, Moira, and Billy

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWELDGEMENTS

I am grateful to my committee members, Brian Taylor, Marty Wachs, Evy Blumenberg, and Joan Walker for their feedback on the initial ideas that formed the basis for this dissertation, and also on earlier drafts of this work. Don Pickrell and Frank Spielberg were both generous in offering feedback on my initial research proposal and sharing the perspectives they had gained through related work that they had completed. Kendra Levine, the librarian for the Institute of Transportation Studies housed at UC Berkeley, was very helpful in helping me track down historical federal reports on transportation funding.

I would also like to thank the thirteen transit professionals who agreed to be interviewed and quoted as part of this research. Their insights and perspectives were invaluable.

This research was funded by the University of California Institute of Transportation Studies, the UCLA Graduate Division, and the University of California Center on Economic Competitiveness in Transportation.

VITA

Carole Turley Voulgaris

4558 7th Avenue NE, Apt 402 (425)502-0019
Seattle, WA 98105 caroleturley@ucla.edu

Education

University of California at Los Angeles Los Angeles, California
Doctor of Philosophy, Urban Planning 2012 – 2017
Expected degree date: June 2017

University of Notre Dame Notre Dame, Indiana
Master of Business Administration, Finance concentration, Magna cum laude 2010 – 2012

Brigham Young University Provo, Utah
Master of Science, Civil Engineering 2006 - 2007

Brigham Young University Provo, Utah
Bachelor of Science, Civil Engineering 2000 - 2006

Research

Ralph, K.M., **C.T. Voulgaris**, & A. Brown. 2017. Travel and the Built Environment: Insights using Activity Densities, the Sprawl Index, and Neighborhood Types. *Transportation Research Record: Journal of the Transportation Research Board,* 2653, 1-9, DOI: 10.3141/2653-01

Ralph, K.M.**, C.T. Voulgaris**, B.D. Taylor, E. Blumenberg, & A. Brown. 2017. Millennials, built form, and travel: Insights from a nationwide typology of U.S. neighborhoods. *Journal of Transport Geography. 57*, 218-226, DOI: 10.1016/j.jtrangeo.2016.10.007

Voulgaris, C.T., B.D. Taylor, E. Blumenberg, A. Brown, & K. Ralph. 2017. Synergistic neighborhood relationships with travel behavior: An analysis of travel in 30,000 US neighborhoods. *Journal of Transportation and Land Use, 10(*1), 437-461, DOI: 10.5198/jtlu.2016.840

Voulgaris, C.T. 2016. A TUF Sell: Transportation Utility Fees as User Fees for Local Roads and Streets. *Journal of Public Works Management and Policy, 21*(4), 305-323. DOI: 10.1177/1087724X16629961

Brown, A., E. Blumenberg, B.D. Taylor, K. Ralph, & **C.T. Voulgaris.** 2016. A Taste for Transit? Analyzing Public Transit Trends among Youth. *Journal of Public Transportation, 19*(1), 49-67. DOI: 10.5038/2375-0901.19.1.4

Voulgaris, C.T., A. Loukaitou-Sideris & B.D. Taylor. 2015. Planning for Pedestrian Flows in Rail Rapid Transit Stations: Lessons from the State of Current Knowledge and Practice. *Journal of Public Transportation, 18*(3), 1-14. DOI: 10.5038%2F2375-0901.18.3.1

Teaching

Urban Planning Math Camp: Instructor Summer 2015
Average overall instructor rating: 4.79 / 5

Urban Planning 256: Transportation Economics and Finance: Special Reader Spring 2015
Average overall instructor rating: 8.56 / 9

Urban Planning 255: Transportation Policy and Planning: Special Reader Fall 2014
Average overall instructor rating: 8.14 / 9

Fellowships and Awards

UC Institute of Transportation Studies Grant 2017
Research funding sponsored by the UC Institute of Transportation
Amount: $20,000

Dissertation Year Fellowship 2016
Fellowship from the UCLA Graduate Division for dissertation research
Amount: $35,829.74

UCCONNECT Dissertation Grant 2016
Grant for dissertation research at campuses in the UCCONNECT consortium
Amount: $7,500

Institute of Transportation Studies Fellowship 2014
Merit-based fellowship for UCLA graduate students in transportation
Amount: $35,582.09

Eisenhower Fellowship 2013
Merit-based fellowship administered by the Federal Highway Administration
Amount: $5,000

Graduate Summer Research Mentorship Grant 2013
Grant from the UCLA Graduate Division for a mentored summer research project
Amount: $6,000

American Public Transportation Foundation Hall of Fame Scholarship 2006
Merit-based scholarship awarded to five transportation students nationally
Amount: $5,000

Service and Leadership

Staffing Working Group 2013-2014
Student committee member

Critical Planning Journal 2012-2013
Managing Co-editor

National Association of Women MBAs 2011
President, Notre Dame Student Chapter

Institute of Transportation Engineers 2006-2007
Secretary, Brigham Young University Student Chapter

American Society of Civil Engineers 2003
Vice President, Brigham Young University Student Chapter

Employment

Transpo Group Kirkland, Washington
Transportation Engineer 2007-2010

Utah Transit Authority Salt Lake City, Utah
Capital Development Intern 2006-2007

Professional Licenses

Professional Engineer (PE), Washington License #: 47256

Chapter 1. Introduction

People make decisions about personal finances, health, education, employment, and interpersonal relationships every day. Firms make decisions about which markets to pursue and where to invest resources. In local and national elections, voters select leaders who make decisions about the sources and allocation of public money and the establishment and enforcement of laws. All these decisions are based on beliefs about the future: both the likely consequences of our decisions and the context in which we will experience those consequences. Throughout history, when the future has seemed particularly uncertain, or when the stakes of a decision have been particularly high, decision-makers have consulted with experts who have (or claim to have) a superior ability to predict the future. In ancient times, such experts might have been oracles and astrologers. In modern times, depending on the specific conditions that interest us, they might be economists, political scientists, meteorologists, or perhaps cost estimators and travel demand modelers. The work of travel demand modelers is particularly important to inform decisions about investments in transit infrastructure, since decision-makers may rely on predictions about future ridership to determine whether a project's benefits are justified by its costs.

Among the objects most closely associated with a supernatural ability to foretell the future is the crystal ball. Today, crystal balls have become a metaphor for this ability. However, in fantasy literature and film, the trustworthiness of crystal ball predictions —and that of the people who make them— varies. Rowling (1999) emphasizes the subjective nature of crystal ball predictions by introducing the character of Sybill Trelawny in her *Harry Potter* series. Trelawny is a divination professor who has genuine (if sporadic) magical abilities to foretell the future. However, she primarily uses a crystal ball to misrepresent the extent of her abilities through vague and unverifiable prophecies. In the film version of *The Wizard of Oz* (Fleming 1939), Dorothy meets Professor Marvel, who pretends to read her fortune in a crystal ball. When she meets Marvel's parallel character —the Wizard— after entering the fantasyland of Oz, his powers prove to be as fraudulent as Professor Marvel's.

Outside of fantasy stories, modern people are less likely to rely on wizards and magicians with crystal balls, choosing instead to turn to experts with sophisticated computer models to foretell the future. But are these modern tools any trustworthier than the uneven performance of crystal balls in fantasy? There is nothing magical about the forecasts made using these modern "crystal balls;" modern forecasting methods are built on plausible assumptions and empirical relationships suggesting how present conditions will likely evolve in the future. But as forecasting models become increasingly data intensive and complex, the process by which they convert input information about current and past conditions into predictions about the future can be far from "crystal-clear." Hence another metaphor has arisen to describe the computer forecasting models that are not well understood even by the technical experts who use them: a black box.

Perhaps even greater than the concern that experts using technical methods to produce forecasts do not really understand the processes embedded in their models is the concern that they *actually do understand* those processes, but that their monopoly on that understanding allows them to manipulate their forecasts to produce whatever result they, or their clients, may wish (Pickrell 1992; Flyvbjerg, Skamris Holm, and Buhl 2005; Richmond 2005; Kain 1990).

This concern is reflected in an old joke about three applicants for the same management position: an accountant, a general civil engineer, and a travel demand modeler. The job interview consists of a few brief questions about each applicant's qualifications, after which the applicant is presented with a column of numbers and asked to sum them up. The first interview is with the accountant. Upon receiving the numbers, the accountant pulls out a calculator, quickly punches them in, and responds, "The total is 512." The second interview is with the general civil engineer, who briefly glances at the numbers and says, “It looks like it’s about 500, give or take.” Finally, the travel demand modeler is brought in for the interview. As with the first two applicants, the interviewer hands the modeler a column of numbers and asks him to sum them up. The modeler looks over his shoulder, leans forward, and asks quietly, “What would you *like* these numbers to add up to?”

No forecast is made in a vacuum. In the case of forecasts prepared for transportation infrastructure projects, the preferences of clients, political leaders, the general public, and project funders may serve to make truly objective analysis elusive if not unattainable. As much as we might like to believe that forecasts come from a magical crystal ball that provides clear, complete, and objective descriptions of future conditions, the forecasting process may more closely resemble a black box into which analysts input a collection of assumptions and imperfect data to produce an output that may tell us more about their hopes or fears than about their reasonable expectations.

In response to concerns about the credibility of cost and ridership forecasts prepared for federal funding applications, the Federal Transit Administration (FTA) has modified its policies over the years to both reduce the reliance on and improve the credibility of such forecasts, particularly for projects that are the subjects of applications for funding through its primary capital grants program, the New Starts program. The purpose of this dissertation is to examine the effect, if any, that federal policy and other changes in the forecasting context have had on the accuracy of cost and ridership forecasts prepared for proposed rapid transit projects applying for funding under the federal New Starts program. To do this the remainder of this dissertation is divided into seven chapters.

Chapter 2 presents a history of the New Starts program and describes how the role of the federal government in funding infrastructure projects for urban transportation was established fairly recently —not until the mid-twentieth century— and was somewhat controversial in its early decades. This controversy was resolved, in part, as procedures and metrics for evaluating projects for funding became standardized. These procedures and metrics have varied over time, but they have always relied on projections of capital costs and ridership to determine the cost-effectiveness of proposed projects.

Chapter 3 summarizes prior literature on forecasting for transit infrastructure, with a particular emphasis on research and theory related to general issues that arise in forecasting across a variety of disciplines, including planning, geopolitics, public health, psychology, and meteorology. Within this literature, forecasts are evaluated based on the methodology and judgment used to produce them, their accuracy with respect to observed values, and their usefulness to decision-makers. Chapter 3 concludes by applying these general principals to a discussion of the literature that specifically address cost estimation and ridership forecasting.

Chapter 4 presents the primary hypothesis for this dissertation: that changes in the New Starts program over the years have served to change perspectives on the purpose of cost and ridership forecasts; that these changes in perspective have in turn led to changes in transit planning and forecasting practice; and that these changes in practice have led to observable changes in forecast accuracy. Chapter 4 goes on to describe the mixed-methods approach I took to testing this hypothesis through interviews with transit professionals and a quantitative analysis of forecasts for projects that were completed using New Starts funds over a period of about fifty years.

Chapter 5 presents the results of the interviews with transit professionals. Based on comments from the interviews, I identify a gradual shift in perspectives over time about the purpose of forecasts for New Starts projects. This shift can be described in terms of three distinct time periods. Before the end of the Reagan Administration in 1989, the perceived purpose of New Starts project evaluation was to limit the role of the federal government in funding local transit projects. Throughout the 1990s and continuing into the early 2000s, perceptions shifted toward an emphasis on fairness and competition among cities competing for scarce federal resources. Eventually, there was a transition to a third era where the focus shifted again from fairness to improving local planning practice.

Chapter 6 describes quantifiable changes in planning practice that have accompanied the changes in perception described in Chapter 5. Over the history of the New Starts program, there have been trends towards shorter construction durations, longer (followed by shorter) horizons for ridership forecasts, and preferences for particular modes at different points in time.

Chapter 7 summarizes the changes in forecast accuracy for initial cost estimates, final cost estimates, and ridership forecasts that have accompanied the changes in perspective described in the Chapter 5 and the changes in practice described in Chapter 6. I find that the accuracy of both ridership forecasts and initial cost estimates has improved substantially over time. The accuracy of final cost estimates has, for the most part, been relatively good and relatively stable over time. For both initial cost estimates and ridership forecasts, later years of cost estimate preparation and shorter construction durations are associated with greater accuracy. Lower-cost projects with higher federal funding shares are likewise associated with more accurate initial and final cost estimates. Finally, forecasts prepared during times of low local unemployment and for light rail projects tend to be the most accurate for both final cost estimates and ridership forecasts.

Chapter 7 also presents the results of a series of regression models that suggest that four variables have an independent relationship with forecast accuracy. Project mode is associated with differences in forecast accuracy for all three forecast types (ridership, initial cost, and final cost). Additionally, shorter construction durations and later forecast preparation years are associated with more accurate ridership forecasts, and lower-cost projects are associated with more accurate initial and final cost estimates.

Chapter 8 concludes this dissertation by relating the findings presented in Chapters 5, 6, and 7 and discussing the degree to which changes in perceptions about the purpose of cost estimation and ridership forecasting for New Starts projects (as described in Chapter 5) may have caused the changes in practice described in Chapter 6, and finally how all of these combined changes may have contributed to improvements in forecast accuracy over time.

Cost and patronage forecasting for new transportation projects has a checkered, and some would say sordid, history over the last half-century. While the capacity and sophistication of the computer models used to produce these forecasts have increased by orders of magnitude over this period, this research suggests that the policies guiding their use have also had important effects on improving their predictive powers. I hope that the results of this study can improve the transparency of the black boxes that produce forecasts for fixed-guideway transit projects, even if the clarity of a crystal ball remains unattainable.

Chapter 2. History of the New Starts program

Today, the Capital Investment Program, more often referred to as the New Starts program, is the Federal Transit Administration's (FTA's) primary grant program for capital investment in fixed-guideway public transit projects, a category that includes urban rail, commuter rail, and bus rapid transit projects. Through the New Starts program, the FTA has allocated over $3 billion per year to urban transit projects over the past twenty years (Federal Transit Administration 1996a, 1997, 1998, 1999, [a] 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, [a] 2011, [a] 2012, [a] 2013, 2014, [a] 2015). However, the federal role in funding public transit capital projects has occasionally been controversial and, partly as a result, has evolved substantially over the past eight decades.

This chapter presents a history of development of the New Starts program, followed by a more detailed discussion of three specific types of changes in the New Starts program over the years: changes in the project development process, changes in project evaluation criteria, and changes in the level and type of project oversight and technical assistance the FTA provides. Table 1 lists the major laws, policies, and rules that I discuss in the following sections.

Development of the New Starts Program

The earliest federal involvement in funding public transit projects was through investments of the Public Works Administration (PWA) in public works projects (Weiner 1999) to combat unemployment caused by the Great Depression. From 1933 to 1935, the PWA invested over $3 billion (over $53 billion in 2017 dollars) in over 30,000 public works projects (J. S. Smith 2006). This included a grant and loan of over $23 million dollars (over $400 million in 2017 dollars) for New York City subway construction (Hood 2004) and smaller investments in few other mass transportation projects around the country (Walker et al. 2016).

Table 1. Legislation, policies, and rules that shaped the New Starts program

Year	Name	Type	Influence on the New Starts program	Influence on forecasting for New Starts
1933	National Industrial Recovery Act	Legislation	• Established the Public Works Administration and set precedent for federal investment in urban transit.	• No formal method for evaluating cost effectiveness (in terms of cost and ridership) as a basis for project selection, although cost estimates would have been required for loans and grants.
1961	Housing Act	Legislation	• Created a loan program for public transit systems.	
1964	Urban Mass Transportation Act	Legislation	• Created the Urban Mass Transportation Administration • Established funding for the New Starts grants program.	
1966	Amended Urban Mass Transportation Act	Legislation	• Extended federal assistance to include the cost of planning, engineering, and design.	
1976	Major Urban Mass Transportation Investments	Statement of Policy	• Formalized evaluation for New Starts projects. • Cost effectiveness as primary selection criterion.	• Required cost and ridership forecasts. • Final funding decisions required earmarks.
1984	Major Urban Mass Transportation Investments	Statement of Policy	• Defined cost effectiveness as cost per new rider (i.e. trip), with both cost and ridership defined relative to a transportation system management (TSM) alternative.	• Clarified thresholds for cost and ridership that forecasts for proposed projects must meet.
1987	Surface Transportation Uniform Relocation Assistance Act (STURAA)	Legislation	• Codified definition and primacy of cost effectiveness as contained in 1984 and 1987 Statements of Policy. • Established the Project Management Oversight (PMO) program to assist project sponsors in to ensure	• Provided assistance to project sponsors in for preparing accurate cost estimates.
1991	Intermodal Surface Transportation	Legislation	• Required the FTA to base project evaluation on mobility improvements, environmental benefits, and	• Reduced the importance of forecasts.
1997	Transportation Equity Act for the 21st Century (TEA-21)	Legislation	• Added Final Design stage to project development process. • Added requirement for Before and After Studies.	• Required intermediate forecasts between selecting alternative and receiving funds. • Required *ex-post* evaluation of forecasts.
2000	Major Capital Investment Projects	Final Rule	• Introduced Transportation System User Benefits (TSUB) to incorporate travel-time savings into benefit calculation. • The TSUB measure was not put into effect until 2003. • TSUB favored projects covering longer distances.	• The 2003 introduction of the Summit software package used to calculate the TSUB measure helped forecasters to uncover many errors within existing travel demand models.
2005	Safe, Accountable, Flexible, Efficient Transportation Equity	Legislation	• Heavy reliance on Congressional earmarks for transportation infrastructure sparks public backlash. • Created Small Starts program.	• Potential for earmarks mitigate consequences of unfavorable forecasts.
2012	Moving Ahead for Progress in the 21st Century (MAP-21)	Legislation	• Eliminated Congressional earmarks. • Created Core Capacity program. • Streamlined project development process. • TSUB replaced with total project ridership. • Eliminated TSM alternative. • Added congestion relief as evaluation criterion.	• Elimination of earmarks increases consequences of unfavorable forecasts. • Shortened duration of forecast horizons. • Summit no longer used to review ridership models. • Forecasting for a TSM alternative not required. • Reduced importance of forecasts.

While these early investments may have set a precedent for federal investment in local public transit, they did not reflect anyone's belief that the federal government should have an ongoing role in funding public transit infrastructure (Weiner 1999).

The first federal legislation to direct funds to local public transit systems for transportation purposes (rather than for job-creation purposes) was the Housing Act of 1961, which introduced a small, low-interest loan program for public transit systems (Weiner 1999). Around that time, President John F. Kennedy requested a joint report on Urban Mass Transportation from the Secretary of Commerce and the Housing and Home Finance Administrator. The report was delivered in March of 1962 and introduced the idea that public transit, as a component of integrated urban transportation systems, was a federal concern (Weiner 1999).

The following month, President Kennedy addressed Congress on the topic of transportation, urging "substantial expansion…in the urban mass transportation program authorized in the Housing Act of 1961" (Kennedy 1963, 300) and the authorization of funds for "economic and technological research in the field of urban mass transportation" (Kennedy 1963, 303) including grants for demonstration projects. Congress responded with the passage of the Urban Mass Transportation Act in 1964, which created the Urban Mass Transportation Administration (UMTA) and authorized federal capital grants for up to two-thirds of the capital costs (construction, reconstruction, and acquisition costs) of public transit facilities (Weiner 1999). In 1966, the Act was amended to extend this assistance to planning, engineering, and design of transit facilities (Weiner 1999).

In 1976, UMTA published a statement of policy acknowledging that, in the years immediately after the passage of the Urban Mass Transportation Act of 1964, most federal money that was directed towards local public transit was "directed toward the preservation of urban transit service in selected cities through the conversion of failing private transit companies to public ownership" (Urban Mass Transportation Administration 1976, 41,512), but that, over the course of a decade, these funding activities were eclipsed by "modernizing existing transit properties and constructing new transit facilities" (Urban Mass Transportation Administration 1976, 41,512). As a result of this shift in focus, the number of cities competing for federal funds was no longer limited to a small number of cities that were already operating rail transit service in the 1960s. As the number of cities applying for federal transit funds increased dramatically, UMTA staff sought to establish a formal process to determine which projects should be prioritized for funding. This process was encapsulated in a 1976 policy statement, effectively marking the beginning of the New Starts program that exists today (Duff et al. 2010).

When Ronald Reagan was elected President in 1980, his transition task force for the Department of Transportation identified four guiding principles for transportation policy: (1) restrict federal involvement to roles that could not filled by the private sector or by state and local government, (2) minimize regulation, (3) make decisions based on cost-benefit analysis, and, (4) fund transportation expenditures with revenues that distribute costs roughly in proportion to benefits (Davis 2015). Taken together, these principles fostered skepticism that the federal government should fund New Starts projects at all. When Reagan approved a five-cent increase in the federal gas tax in 1982, he was concerned about being "trapped into funding mass transit projects in the future" (Davis 2015, 25) and specified that fuel tax revenues could not be used for funding New Starts projects (which were and continue to be funded from general funds, rather than from the Highway Trust Fund) (Davis 2015). This hesitancy toward funding new urban rail projects was supported by early case studies suggesting that such projects were generally not a cost-effective use of federal money, since federal aid had the potential to distort local decision-making in favor of projects that would not otherwise be financially feasible (Gomez-Ibanez 1985a, [b] 1985).

In contrast to the skepticism the White House had developed toward funding transit infrastructure, members of Congress often were eager to use earmarks to fund projects in their home districts. In a 1987 report to Congress, UMTA described how Congressional earmarks hampered efforts to select projects based on merit.

> Local interests persuade their representatives in Washington to “earmark” Federal discretionary capital funds for the local projects. The investment decision is often not predicated on the merits of the project nor on the ultimate cost of the projects earmarked. Rather, it is based on political expediency. … In an effort to assure that discretionary capital grants would only go to the best transit investments, UMTA adopted its Major Capital Investment Policy in May of 1984. … Unfortunately, Congress has not supported the UMTA Major Capital Investment Policy, and has proceeded to earmark the discretionary funds available for projects without regard to their actual benefits. … The ultimate result of transit decisionmaking based on grantsmanship rather than merit has been the failure of transit capital projects to achieve the benefits claimed for them. The only way sound transit projects will be developed is to take away the attraction of federal discretionary grants and insist on comprehensive financial planning. (Urban Mass Transportation Administration 1987, 145–151)

In spite of these objections, the practice of earmarking projects to bypass recommendations by UMTA —or its successor, FTA— to review and select projects based on merit continued for decades. For example, in the 2007 fiscal year, of the 66 projects that received New Starts funding, 32 were funded based on Congressional earmarks alone, rather than on FTA’s evaluation and ratings process (Office of Inspector General 2007).

Public awareness of and concern about Congressional abuse of earmarks for transportation projects reached an apex in the early 2000s. In addition to a dozens of earmarks for New Starts transit projects, the 2005 transportation reauthorization bill —Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users, or SAFETEA-LU— also included many other transportation earmarks, including one for a bridge from the small city of Ketchikan to largely unpopulated Gravina Island in Alaska that cost taxpayers an equivalent of $40,000 per resident of Ketchikan (M. A. Boyle and Matheson 2009; 109th Congress 2005). Nicknamed the “bridge to nowhere” this project caught media and public attention as a symbol of wasteful federal pork barrel spending. Subsequent partisan finger-pointing over the project precipitated bipartisan efforts toward earmark reform beginning in 2006 (Doyle 2011; M. A. Boyle and Matheson 2009). As a result of this attention and these efforts, the subsequent transportation reauthorization bills included *no* earmarks for highway or transit projects (112th Congress 2012; 114th Congress 2015), so all subsequent New Starts grants were based on the FTA's ratings and evaluation process — a remarkable turnabout from decades of political practice.

Another important contribution of SAFETEA-LU to the evolution of the New Starts program was the creation of the complementary Small Starts program. Under Small Starts, projects that were anticipated to cost less than $250 million and request less than $75 million in federal assistance are subject to a streamlined process for project development and evaluation (109th Congress 2005). The Small Starts program has been carried through to subsequent transportation reauthorization bills. Under the current bill —the Fixing America's Surface Transportation Act, or the FAST Act, of 2015— Small Starts projects are those with total project costs of less than $300 million requesting less that $100 million in federal assistance (114th Congress 2015).

Moving Ahead for Progress in the 21st century (MAP-21), the 2012 federal surface transportation legislation, added a third program (in addition to New Starts and Small Starts) under the umbrella of the Capital Investment program: the Core Capacity program. Projects can qualify for funding under the Core Capacity program if they increase the capacity of an existing transit corridor by at least ten percent and demand in the corridor is anticipated to exceed capacity (without the project) within the next five years (112th Congress 2012). This change only affected a small number of cities with fixed-guideway transit corridors that were operating at or near capacity. A much more far-reaching change by MAP-21 was to streamline the project development process for New Starts.

The New Starts Project Development Process

UMTA's 1976 policy statement (Urban Mass Transportation Administration 1976) introduced the New Starts project development process. Under that policy, applicants for New Starts funding —who are referred to as "project sponsors" and are typically local transit authorities— would begin the project development process by defining a set of alternative projects that could meet a local need, developing a plan for public involvement in selecting an alternative, and establishing a methodology for comparing alternatives (including a demand forecasting methodology). After UMTA approved the alternatives and the methodologies for evaluating them, the project sponsor would complete the Alternatives Analysis (AA) concurrently with the Draft Environmental Impact Statement (DEIS) required by the National Environmental Protection Act (NEPA) for all federal agency actions (such as funding New Starts projects). After UMTA approved the completed AA study, a project would be eligible to advance to the next stage in the project development process: Preliminary Engineering (PE). The PE stage would be concurrent with the preparation of the Final Environmental Impact Statement (FEIS) required by NEPA. During the PE stage, a project sponsor would also obtain evidence of revenue sources for the local share of a project's funding and evidence of local support for the project. Upon completion of PE, a project sponsor could apply for a capital grant for final engineering and construction.

Two years later, UMTA released a second policy statement clarifying and reemphasizing the criteria for evaluating projects and introducing a multi-year commitment of federal funds through a Letter of Intent (later called a Full-Funding Grant Agreement or FFGA) that could be awarded to projects upon completion of PE (Urban Mass Transportation Administration 1978). These multi-year funding commitments reduce the risk that project sponsors will begin a project with an expectation of federal funding, only to have Congress decline to continue appropriating expected funds until the project is completed.

The next major change to the project development process came with the passage of the 1997 transportation reauthorization bill —the Transportation Equity Act for the 21st century (TEA-21)— which added the Final Design (FD) stage between the completion of PE and the award of an FFGA (Duff et al. 2010). With this change, project sponsors were required to obtain FTA approval before advancing from PE to FD, and were not eligible for an FFGA (which would previously have been awarded upon completion of PE) until the completion of FD. TEA-21 and the subsequent Final Rule issued by the FTA (2000c) also made an important change to the New Starts process by introducing a requirement for *ex post* analysis through a Before and After Study of completed New Starts projects. This latter requirement figures importantly in this dissertation, as discussed in Chapter 4. The purpose of the *ex-post* analysis is two-fold: first, to evaluate the accuracy of cost and ridership forecasts, and second, to determine the actual effects of the project.

Short and Kopp (2005) describe transportation planning processes in Europe and argue that project appraisal is a consistent weakness there, noting that, "*ex ante* appraisal is often biased and *ex post* appraisal rarely takes place." While this observation refers to processes in Europe, similar shortcomings can be observed in the United States. Why is there reluctance to perform *ex post* analyses? *Ex post* analysis requires time and resources that project sponsors may prefer to allocate to other activities that offer more clear and immediate benefits. Project sponsors may also wish to protect themselves from findings that could make their earlier forecasts, if they prove significantly inaccurate, appear to be the products of either intentional deception or incompetence. By forgoing such analyses, they may avoid being publicly shamed by prior forecasts, but they likely miss an opportunity to learn from them as well. Thus, opportunities to improve their techniques (and/or tone down their optimism) for future projects may be lost.

The 2000 Final Rule, "Major Capital Investment Projects," describes the comments received by the FTA when this requirement was initially proposed. Two commenters —one interest group and one trade association— addressed the time period that should be covered by Before and After Studies. Both were supportive of the idea of *ex post* project evaluation, but neither supported the use of opening-year data for this purpose (Federal Transit Administration 2000c).

> The interest group recommended that reviews not occur until at least after the first year of revenue service, and not later than 15 years, suggesting ratings at 2 and 7 years. The trade association recommended that … ratings should encompass a 5-10 year operating period. (Federal Transit Administration 2000c, 76,866)

These comments echo the objections that Flyvbjerg (2005) reports in his earlier work (Flyvbjerg, Skamris Holm, and Buhl 2005; Skamris and Flyvbjerg 1996) evaluating demand forecasts for public transit projects based on opening-year ridership data. He summarizes the argument behind these objections thus:

> If projects experience start-up problems, … then this may initially affect traffic negatively, but it would only be temporarily and it would be misleading to measure inaccuracy of forecasts on that basis, according to this argument. When start-up problems are over, normal operations will ensue, traffic will increase, and this should be the basis on which inaccuracy is measured, the argument continues. Furthermore, it takes time before travelers effectively discover and make use of a new transportation facility and change their travel behavior accordingly. Inertia is a factor. A project with lower-than-forecasted traffic during the first year of operations may well catch up with the forecast a few years down the line and it would be more appropriate to measure inaccuracy on that basis. If the first year of operations is used as the basis for comparison, the result would be the identification of too many underperforming projects, or so the opponents to using this basis argue. … Some forecasts assume so-called demand "ramp up" over the first 3–5 years of a scheme up to the full modeled demand. …By comparing opening years one may thus be assessing two sources of uncertainty: the inaccuracy of the modeling itself, and the inaccuracy of the assumed ramp up to full demand. (Flyvbjerg, Skamris Holm, and Buhl 2005, 525)

While Flyvbjerg (2005) acknowledges that the use of a later year might be ideal to account for the possibility of ramp-up, he argues that this would make it impossible to include a large number of projects in a single study since, although most projects do make opening-year forecasts and collect opening year ridership/traffic data, there is not a common near-term horizon year to allow for comparison across a large number of projects. Even using opening-year data, the sample used by Flyvbjerg and colleagues (which was the largest of its kind and spanned three decades) to study inaccuracy in cost (Flyvbjerg, Holm, and Buhl 2002) and ridership/traffic (Flyvbjerg, Skamris Holm, and Buhl 2005) projections for public works projects included only 27 public transit projects (including urban rail, high speed rail, and conventional rail), with the rest of the 210-project sample comprising road projects.

Although both researchers and practitioners have acknowledged the possibility of ridership ramp-up, the existence of this phenomenon has not been empirically studied, and its extent and magnitude have not been systematically documented.

In response to concerns with using opening-year data to evaluate the accuracy of forecasts, the FTA requires that comparisons between forecasts and observed data be based on observations taken after two years of revenue service (Federal Transit Administration 2000c). They explain,

> FTA recognizes that this evaluation will provide only a short-term "snapshot" of the performance of a new fixed-guideway system, and that many of the benefits, particularly in terms of land use, are long-term in nature. Project sponsors are of course encouraged to continue their data collection efforts beyond the period two years after opening. However, given the nature of the appropriations and authorization process, there is also a need for short-term data to provide an initial indication of the benefits of a project. (FTA, 2000, 76866)

During the Final Design phase of project development, the project sponsor must submit a plan to the FTA that provides for the collection and analysis of data on project costs and performance after project completion (Duff et al. 2010). Throughout the project development process, sponsors must also maintain a record of assumptions, data, and methodologies that were used to create the forecasts at each phase. Completed Before and After studies compare both the assumptions used as inputs to forecasts and the forecasts themselves to observed conditions two years after project completion. In instances where the design and/or scope of the project has changed between the time the forecast was made and the time the project was completed, the Before and After Study will typically present the original forecast as well as a version of the original forecast with adjustments for changes in project scope.

With the introduction of an additional phase in the project development process and the requirement to produce Before and After Studies for completed projects, TEA-21 increased the complexity of the process for applying for New Starts funds. Fifteen years later, MAP-21 mitigated this added complexity by streamlining the project development process, while retaining the requirement for Before and After Studies. While an Alternatives Analysis is still required under NEPA, it is no longer a stage in the New Starts project development process. Instead, the first stage in the project development process is known as "project development" (Emerson 2012). No evaluation or rating by the FTA is necessary to enter the project development stage; projects are accepted in this phase when a project sponsor submits a request and a project description (Emerson 2012). The NEPA process is completed during the project development stage, after which the FTA rates and evaluates the project for entry into the "engineering" stage, which is roughly equivalent to the former FD stage. Thus, MAP-21 reduces the number of milestones at which the FTA must evaluate a project from four to two. The project development phase is also limited to a duration of no more than two years, although the FTA may grant extensions as necessary (Emerson 2012).

Project Evaluation Criteria

While there have been only a few changes to the New Starts process since the mid-1970s, there have been many changes to the criteria that UMTA and later FTA used to evaluate potential projects. Cost effectiveness (cost per unit of benefit) has been a consistent evaluation criterion through the history of the New Starts program, but the method of calculating project benefits has undergone important changes. Moreover, several additional criteria have been added over the years.

Calculating cost effectiveness

The 1976 policy statement specified that projects should be evaluated based on their anticipated cost effectiveness, but did not offer specifics about how project benefits would be measured for inclusion in a cost-effectiveness calculation. In fact, UMTA staff seemed to go out of their way not to commit to any specific benefits measure.

> Since each metropolitan area has differing characteristics, Federal mass transportation assistance cannot be based on standardized prescriptions… The Federal Government does, however, have a strong interest in ensuring that Federal funds available for mass transportation investments be used …with maximum effectiveness….Federal support will be available only for those alternatives which … analysis has demonstrated to be cost-effective, where effectiveness is measured by the degree to which an alternative meets the locality's transportation needs, promotes its social, economic, environmental and urban development goals, and supports national aims and objectives (Urban Mass Transportation Administration 1976, 41,513).

Thus, cost-effectiveness was established as a primary selection criterion, but remained somewhat broadly defined. This changed in 1984 with a policy statement that introduced a ratings methodology to compare competing projects (Urban Mass Transportation Administration 1984). The ratings system introduced in the 1984 policy statement defined cost effectiveness as the cost per new rider, where both the numerator (cost) and the denominator (ridership) were to be calculated relative to a transportation system management (TSM) alternative, defined as a low-cost improvement to bus operations. Thus, high-cost projects could be justified only by high ridership forecasts. While the 1984 policy statement acknowledges that transit projects have many secondary benefits in addition to those that accrue to riders, it justifies the emphasis on ridership:

> It is … an accurate index of many of transit's potential secondary benefits, including the structuring of urban development patterns and reductions in congestion, pollutant emissions and energy consumption (Urban Mass Transportation Administration 1984, 21,286).

In 1987, Congress codified cost effectiveness as the primary project evaluation criterion in the Surface Transportation and Uniform Relocation Assistance Act (STURAA), an authorization act for Federal programs for surface transportation.

One criticism of measuring project benefits in terms of the number of new riders attracted to a project was that this measurement did not account for benefits a project might offer to existing riders, particularly reductions in travel time resulting from replacing conventional bus service fixed-guideway service. In an attempt to properly account for all project benefits, FTA introduced a new measure of project benefits in 2000, referred to as Transportation System User Benefits (TSUB), which included travel-time savings for existing riders in addition to benefits to new riders (Federal Transit Administration 2000c).This new measure did not become effective until 2003 (General Accounting Office 2003), when the Summit software package was introduced, which I discuss further later in this chapter.

In general, project sponsors appreciated this more complete measure of project benefits, but found its complex calculation to be unintuitive and difficult to calculate and communicate to project stakeholders (General Accounting Office 2003; Federal Transit Administration 2013c). In connection with the MAP-21 efforts to streamline project evaluation, FTA changed the calculation for cost-effectiveness again in 2013. Project benefits were defined in terms of the total number of trips forecast to occur on the new project, and both costs and benefits were included in the calculation as total values rather than as incremental values relative to a TSM alternative (Federal Transit Administration 2013c). As discussed in Chapters 5 and 6, these changes in cost effectiveness had major implications for the types of projects that would be eligible for New Starts funding.

Additional project evaluation criteria

When FTA began to define project benefits strictly in terms of ridership in 1984, many were concerned that secondary project benefits that were not directly related to ridership would not be accounted for in project evaluation (although Pickrell (1992) has argued that benefits that are truly unrelated to ridership are unrelated to transportation and should thus not be considered by the FTA). In response to these concerns, Congress included in the 1991 transportation reauthorization bill —the Intermodal Surface Transportation Efficiency Act, or ISTEA— a requirement that the FTA also consider additional criteria including mobility improvements, environmental benefits, and operating efficiencies (Duff et al. 2010). In 1996, FTA published a notice implementing specific measures to evaluate New Starts projects based on these new criteria (Duff et al. 2010). The number of measures for project justification increased from one (cost effectiveness) to eleven (Federal Transit Administration 1996b):

- Mobility improvement
    1. Value of projected travel time savings
    2. Number of low-income households living within half a mile of project boarding points
- Environmental benefit
    3. Value of the projected reduction in pollutant and greenhouse gas emissions
    4. Change per year in regional energy consumption
    5. Current and projected compliance with Environmental Protection Agency (EPA) air quality standards
- Operating efficiencies
    6. Forecast change in operating cost per passenger mile
- Cost effectiveness
    7. Incremental cost (capital and operating cost) per incremental rider
- Transit-supportive existing land-use policies and future patterns
    8. Combined rating of low, medium, or high, based on
        - Existing land use
        - Containment of sprawl
        - Transit-supportive corridor policies
        - Performance of land-use policies
- Other factors
    9. The degree to which institutions are already in place as assumed in forecasts
    10. Project management capability
    11. Additional relevant factors

Of the eleven measurements listed above, a few, such as project management capability and the number of low-income families living within a half mile of project boarding points, do not depend on ridership forecasts or cost estimates. Others, such as anticipated cost-effectiveness and reductions in greenhouse gas emissions would vary directly with anticipated ridership and/or cost.

Project justification criteria remained unchanged with the passage of the next highway authorization act, the Transportation Equity Act for the 21st century (TEA-21). However, in a 2005 “Dear Colleague” letter, the FTA re-emphasized the importance of cost-effectiveness, and thence ridership, as a project evaluation criterion by announcing that projects that failed to achieve a medium rating or higher for cost effectiveness would not be recommended for funding (Dorn 2005). In the same year, the subsequent federal surface transportation act, Safe, Affordable, Flexible, Efficient Transportation Equity Act, A Legacy for Users (SAFETEA-LU) added two additional project justification criteria for New Starts projects: economic development and land use (Duff et al. 2010). These additions were presumably intended to reduce the importance of cost-effectiveness in project evaluation (Chapter 5 discusses the actual and/or perceived motivations behind changes in the New Starts program in more detail). However, in 2007, FTA published a notice of proposed rulemaking (NPRM) wherein cost-effectiveness would be assigned a 50 percent weight in the overall project rating, and to continue to require that projects with less than a Medium rating for cost effectiveness not be recommended for funding. Later that year, Congress responded by enacting legislation to prevent implementing these proposed rules (Duff et al. 2010). The following year, Congress passed the SAFETEA-LU Technical Corrections Act of 2008, which required the FTA “give comparable, but not necessarily equal weight to each project justification criteria in calculating the overall project rating” (quoted in Duff et al., 2010). The FTA responded in 2009 with an NPRM proposing that cost-effectiveness be weighted at 20 percent of the overall project rating, and each of the other criteria would be rated at either 10 or 20 percent (Duff et al. 2010).

This back-and-forth reflects a fundamental tension between Congress and the FTA. A majority of members of Congress appear intent on maximizing the opportunity for projects located in their own districts to receive Federal funding, despite the fact that few Congressional districts have the population density and other necessary conditions to achieve ridership levels that would justify Federal investment. Conversely, the FTA leadership, particularly under Republican presidents, appears to favor awarding New Starts funding based on project cost and patronage merits, rather than on local project benefits less directly related to transit use. I discuss these differences more fully in Chapter 5.

In 2012, MAP-21 added congestion relief as an additional project evaluation criterion, for a total of six criteria: mobility improvement, economic development, environmental benefits, congestion relief, cost effectiveness, and land use. In the 2013 Final Rule (Federal Transit Administration 2013c) FTA announced that each of these six criteria would be weighted equally in the overall project rating.

Technical Assistance and Project Oversight

Over the years, concerns about the accuracy of ridership forecasts and cost estimates prepared during the New Starts project development process have prompted FTA to increase its role in technical assistance and project oversight for New Starts projects. The first step in this direction was the establishment of the Project Management Oversight (PMO) program authorized by the 1987 surface transportation reauthorization bill (the Surface Transportation and Uniform Relocation Assistance Act, or STURAA) (General Accounting Office 2000). Under the PMO program, FTA hires a private engineering firm as a project management oversight consultant (PMOC) to work with each project sponsor throughout the project development process to develop a reasonable project budget and schedule, and continues to monitor to the project sponsor to ensure that it stays within the budget and in schedule (General Accounting Office 2000). In a 1997 review of the Los Angeles subway project, FTA learned that the local project sponsors did not have adequate financial capacity to fund its agreed-upon local share for the project. As a result, FTA expanded the role of the PMO program to include an analysis of project sponsors' financial capacity to fund planned local funding shares (General Accounting Office 2000).

The PMO program increased FTA's role in seeking to maximize the accuracy of cost estimates, but these changes did not extend to oversight of ridership forecasting. Beginning with the 1976 policy statement that introduced the New Starts process (Urban Mass Transportation Administration 1976), UMTA/FTA was tasked with approving project sponsors' ridership forecasting methodology, but given the complexity of most regional travel demand models, detailed model review was generally not possible.

In 2003, FTA introduced the Summit software package to assist project sponsors in calculating the TSUB measure of project benefits (General Accounting Office 2005). Although the purpose of the software was to calculate project benefits, it also served as a much-needed quality control for travel demand model specifications, since it required users to enter such detailed model outputs. In testimony before Congress in 2004, the Inspector General of the Department of Transportation described the Summit software as "an important step … to help identify problems with ridership forecasts" (Mead 2004, 14).

When ridership models are too complex to allow for adequate oversight and quality control, the use of a simpler model could address this problem. In addition to the MAP-21 efforts to simplify the New Starts process and the cost effectiveness measure, FTA also took steps at that time simplify ridership forecasting. In the 2013 Final Rule (Federal Transit Administration 2013c) replacing the TSUB measure with total trips on project, FTA also promised to develop a simplified method for project sponsors to calculate trips on project and reductions in vehicle miles traveled (VMT), required for calculating environmental benefits. Later that year, FTA introduced the Simplified Trips-on-Project Software (STOPS) (Federal Transit Administration 2013d).

STOPS is a simplified four-step travel demand model that substitutes the trip-generation and trip-distribution steps with worker flows from the Census Transportation Planning Package (CTPP), applying factors and information about transit trip generators to account for non-work trips and non-home-based trips (Federal Transit Administration 2013d). STOPS uses General Transit Feed Specification (GTFS) data to represent the transit network (Federal Transit Administration 2013d). Rather than representing the roadway network in the model, STOPS simply uses zone-to-zone travel times and distances (Federal Transit Administration 2013d). The STOPS model has been calibrated and validated against data from a variety of fixed guideway systems in metropolitan areas across the country, but allows minimal adjustments for calibration to local conditions (Federal Transit Administration 2013d). At this time, project sponsors may choose to produce their forecasts using either the STOPS model or a locally maintained regional travel demand model that is subject to review and approval by FTA.

Conclusion

Throughout the history of the New Starts program, we can observe a tension between, on the one hand, improving the opportunities for all project sponsors to qualify funding and, on the other hand, creating a system of merit-based competition that would award funds only to a few high-performing cities. Likewise, we see a tension between rigor and simplicity or transparency in project evaluation, and between limiting and expanding the federal role funding and overseeing transit project development and construction. Finally, we can observe a tension among competing (if not necessarily conflicting) priorities such as attracting new transit riders, serving existing riders, and using transit infrastructure to shape urban form and urban economies more broadly.

The changes that have arisen through these tensions have influenced the context in which cost estimates and ridership forecasts are made for New Starts projects. How have these contextual changes affected the ways in which project sponsors perceive the role of forecasting and cost estimating in transit planning? Has forecast accuracy changed as policies and perceptions have changed? A review of the existing literature on forecasting in general — as well as prior research that specifically addresses cost estimating and ridership forecasting for transit— offers some insight into the types of changes we might expect to see in connection with the shifting policy landscape, and this is the focus of the next chapter.

Chapter 3. Literature Review

This chapter discusses literature related to forecasting for public transit infrastructure projects. The bulk of the chapter summarizes theories about and research on forecasting in general, drawing on the work of scholars from diverse disciplines including planning, politics, public health, psychology, and meteorology. Two brief sections follow, which discuss research on cost estimation for capital construction projects and demand forecasting for transportation infrastructure projects. The final section summarizes a small group of studies with scopes similar to this dissertation: cost and ridership forecasts for federally funded transit capital projects in the United States.

Literature on Forecasting

The degree of trust we should place in the forecasts experts produce might vary somewhat by topical area, but studies on forecasting in general provide a helpful starting point for thinking about our reliance on expert forecasts.

Table 2: Categories of criteria for evaluating forecasts

Publication	Types of forecast addressed	Proposed criteria for forecast evaluation		
		Methodology and judgment	**Accuracy**	**Usefulness**
Ascher, William. 1979. Forecasting: An Appraisal for Policy-Makers and Planners. Baltimore: Johns Hopkins University Press.	Population, economic, energy, transportation, and technology	Insider's approach	Outsider's approach	Discussed, but not an evaluation criterion
Tetlock, Philip E. 2009. Expert Political Judgment: How Good Is It? How Can We Know? Kindle Edition. Princeton University Press.	Political events	Logical coherence and process tests	Empirical correspondence tests	Discussed, but not an evaluation criterion
Murphy, Allan H. 1993. "What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting." Weather and Forecasting 8 (2): 281–93.	Weather	Consistency	Quality	Value
Kahneman, Daniel. 2013. Thinking, Fast and Slow. Reprint edition. New York: Farrar, Straus and Giroux.	Project Planning	Inside view	Outside view	Not discussed

How should we go about evaluating forecasts prepared by experts? Three scholars, Ascher (1979), Tetlock (2009), and Murphy (1993), have proposed related categories of forecast-evaluation criteria. Kahneman (2013) describes two approaches to preparing forecasts that relate closely to the evaluation categories described by Ascher (1979), Tetlock (2009), and Murphy (1993). Table 2 summarizes these criteria; the terms used in this table are discussed below.

In introducing his evaluation of population, energy, economic, transportation, and technology forecasts, Ascher (1979) differentiates the "insider's approach" from the "outsider's approach." The insider's approach focuses on the scientific validity of the forecasting technique and the correct use of the information that was available to the forecaster at the time the forecast was made. In contrast, the "outsider's approach" focuses on the accuracy of the forecasts rather than the method that is used to produce them.

Tetlock (2009) likewise proposes two categories of tests to evaluate a political forecaster's judgment: "correspondence tests rooted in empiricism" (Tetlock 2009, 7), which correspond neatly with Ascher's (1979) outsider's approach, and "Coherence and process tests rooted in logic" (Tetlock 2009, 7), which are closely related to, if somewhat distinct from, Ascher's (1979) insider's approach and address the question of whether the forecast is based on a set of internally consistent beliefs that the forecaster updates in response to new evidence.

In reference to weather forecasts, Murphy (1993) identifies three distinct types of forecast goodness: *consistency*, *quality*, and *value*. *Consistency* is the degree to which a published forecast corresponds to the forecaster's best judgment, and is closely related to (but again somewhat distinct from each of) Ascher's (1979) insider's approach and Tetlock's (2009) coherence and process tests. *Quality* is the degree to which a published forecast corresponds to subsequently observed conditions. This corresponds neatly with Ascher's

(1979) outsider's approach and Tetlock's (2009) correspondence tests, and is typically what we refer to when we discuss forecast accuracy. Murphy's (1993) final type of forecast goodness, *value*, refers to the usefulness of a forecast to the forecast's users. While Ascher (1979) and Tetlock (2009) both discuss the concept of forecast usefulness, neither suggests it as a separate criterion for evaluating a forecast.

Kahneman (2013) does not directly address how a forecast should be evaluated, but rather discusses two approaches to *preparing* project-management forecasts: the "inside view" and the "outside view." A forecast prepared with an inside view is based on the forecaster's knowledge of specific characteristics of the planned project. The inside view is closely related to forecast evaluation based on Ascher's (1979) insider's approach, Tetlock's (2009) coherence and process tests, and Murphy's (1993) consistency. A forecast prepared with an outside view is based on the observed data from completed projects with similar characteristics to the planned project. The outside view is closely related to Ascher's (1979) outsider's approach, Tetlock's (2009) correspondence tests, and Murphy's (1993) quality.

Taken together, the loosely related frameworks these four scholars propose suggest three dimensions that are useful in evaluating forecasts: (1) methodology and judgment, (2) accuracy, and (3) usefulness. I review each of these dimensions in greater detail below, followed by a discussion of the relationships among them.

Methodology and judgment

Forecast methodology comprises three components: (1) data on historical or existing conditions, (2) a theoretical and/or mathematical model, and (3) the forecaster's judgment.

In evaluating many forecasts from a variety of disciplines, Ascher (1979) finds that improvements in the accuracy of forecast inputs (which are often also forecasts) have a much greater influence on forecast accuracy than improvements in the theoretical/mathematical model used to produce the forecast. Alonso (1968) demonstrates how model improvements may be especially ineffective when they increase model complexity, since complex models tend to magnify the uncertainty of model inputs. As a result, when input data are uncertain (either due measurement error of inputs that refer to existing conditions, or forecasting error in inputs that refer to future conditions), a less conceptually complete, but simpler, model will yield a more reliable forecast than a more conceptually complete, more complex model. This effect of model complexity may help explain Ascher's curious finding that the introduction of sophisticated modeling techniques in a variety of forecasting disciplines has generally not improved forecast accuracy (Ascher 1979), since any improvement in the quality of input data is likely to be offset by the magnification of uncertainty that usually accompanies increasing model complexity. Nevertheless, as computers have become more powerful and widespread, complex mathematical models have played a greater role in forecasting.

Innovation diffusion theories are useful in understanding why forecasters may choose to adopt new forecasting approaches. Rogers (1962; 2003) was the first to propose a general theory of innovation diffusion to apply across all types of innovation. He defines diffusion in terms of four elements as "the process by which (1) an *innovation* (2) is *communicated* through certain *channels* (3) *over time* (4) among member of a *social system*" (Rogers 2003, 11). The adoption of a technology or innovation thus a function of the characteristics of the innovation itself, the ways in which those characteristics are communicated, the characteristics of the communications channels, the passage of time, and the characteristics of the social system of potential users.

Rogers (2003; 1962) goes on to identify five perceived attributes of innovations that help to explain their adoption: (1) *relative advantage* over the ideas the innovation is intended to replace; (2) *compatibility* with the values and needs of potential adopters; (3) *complexity*, where innovations that are difficult to understand are less likely to be adopted;

(4) *trialability*, or the degree to which potential adopters can experiment with the innovation on a limited basis; and (5) *observability*, or the degree to which the benefits of an innovation can be observed by those who have not yet adopted it.

Based on the framework described above, the very complexity of increasingly sophisticated forecasting models may impede their adoption, especially when their relative advantage over simpler models is not at all clear — as is often the case for innovations in forecasting. Spyros Makridakis, a leading scholar on forecasting, has organized a series of three competitions referred to in the forecasting literature as the Madridakis competitions or M competitions (Makridakis and Hibon 2000; Makridakis et al. 1982, 1993; Hyndman and Koehler 2006). The purpose of these competitions has been to compare the performance of a variety of extrapolative methods for forecasting time series data. Two findings that were consistent in all three competitions are relevant in evaluating forecast methodologies. First, more complex methods generally do not perform better than simpler methods. Second, the results obtained by averaging forecasts produced using a variety of methods tend to be more accurate than the forecasts produced by any individual method included in the average (Makridakis and Hibon 2000; Makridakis et al. 1982, 1993). Interestingly, the latter finding is consistent with Tetlock's (2009) observation about political forecasts prepared based on expert judgment alone: that averaging predictions by a group of experts will generally yield a more accurate set of predictions than any one expert will produce.

These findings would suggest that forecasting resources would be better allocated to applying a variety of simple models and averaging or combining the results than to developing a single complex model. However, the increasing availability and affordability of computing power has served to sharply reduce the cost of increasing model complexity. As forecasting models have become more complex and reliant on high-power computing, some have expressed concern that these models are increasingly becoming "black boxes": mysterious machines that transform inputs into outputs using methods that even the forecaster operating the machine may not fully understand (Ascher 1979; Silver 2015). An argument could be made that one advantage of a black-box model is that it reduces reliance on the modeler's subjective and biased judgment. Fair (1971) argues that if the modeler has the ability to tinker with model parameters and inputs to achieve his preferred result, the model offers little advantage over the modeler's subjective judgment. Ascher (1979) and Silver (2015) both describe this concern with an analogy to the Mechanical Turk, the subject of an essay by Edgar Allen Poe about a machine that astonished audiences throughout Europe in the late 18th and early 19th centuries by its ability to apparently play chess independently against a human opponent. Poe correctly surmised that the machine was actually operated by a concealed human chess master (Poe 1836). Just as the chess master's abilities were more impressive when it appeared that they were those of a machine, is it possible that the primary function of computer models is to add credibility to the human modeler's subjective opinions?

There is some evidence that good judgment on the part of a human model operator improves model performance. Ascher cites "[n]umerous studies [which] have shown that, contrary to expectations, short-term economic forecasts that require the forecaster to project the values for exogenous variables often are much more accurate than parallel forecasts that incorporate the *actual* values for these exogenous variables" (Ascher 1979, 68, emphasis in orginal). This occurs because the forecaster iteratively adjusts model inputs to bring the forecasts into a range that is consistent with the forecaster's judgment, which can compensate to some degree for errors or omissions in the model specification, but may obscure serious weaknesses in the validity of the model. This finding does not necessarily mean that the model does not add value, or that we would do equally well to rely on expert judgment alone.

> The model serves primarily to assess judgmentally the general implications of the forecaster's assumptions on future exogenous developments, including his ad hoc adjustments for anticipated changes in structure since the sample period, and for the correction of apparent specification errors. *(Hickman 1972, 17)*

Ultimately, the question of whether a forecaster should exercise human judgment in preparing a forecast is moot; even the most mechanistic models require human judgment in selecting input variables and presenting model results, to say nothing of determining whether the model is appropriate for addressing the relevant question in the first place. Silver (2015) warns of the danger of failing to exercise judgment in working with computer models:

> Be wary… when you come across phrases like "the computer thinks…" [I]f you get the sense that the forecaster means this … literally—that he thinks of the computer as a sentient being or the model as having a mind of its own—it may be a sign that there isn't much thinking going on at all. Whatever biases and blindspots the forecaster has are sure to be replicated in his computer program (Silver 2015, 293).

Murphy (1993) specifies that, to achieve *consistency*, a forecast should be *consistent* with the forecaster's judgment in two respects: first, with respect to the estimate itself, and second – since no statement about the future can be made with 100 percent certainty – with respect to the level of certainty associated with the estimate. Murphy (1993) goes on to point out that *consistency* can be difficult to ascertain by outside observers, since only the forecaster herself can determine what her own best judgment was and check it for consistency against her published forecast. However, Tetlock (2009) argues that forecasters should be held to "coherence and process tests rooted in logic. Are their beliefs internally consistent? And do they update those beliefs in response to evidence?" (Tetlock, 2009, 348-349). Such tests refer to something subtly different from Murphy's (1993) *consistency* in the sense that they do not ask whether a forecaster's beliefs are consistent with her forecasts, but rather whether she is justified in holding those beliefs, sincere as they may be, based on the evidence available to her.

A forecaster's judgment of whether a forecast is within a reasonable range may be somewhat subjective, but the forecaster will ideally update his subjective beliefs based on data. Kahneman (2013) describes how project performance forecasts prepared based on the characteristics of the project itself (described as the "inside view") can be improved when the forecaster incorporates information about the performance of similar projects that have already been completed (the "outside view"). Thus the forecaster's judgment about whether the initial forecast is reasonable is informed by data on similar projects. Kahneman and Tversky (1979) suggest reference class forecasting as means of going beyond simply anchoring the forecaster's judgment based on prior performance and applying a formal, quantitative technique to actually develop forecasts based on probability distributions derived from data on other projects. This method was first operationalized and applied to project planning by Flyvbjerg (2006).

Tetlock (2009) and Silver (2015) both discuss how a forecaster should update beliefs according to new evidence by applying Bayes' theorem. Based on the work of Reverend Thomas Bayes (Bayes and Price 1763) and further refined by Pierre Simon, Marquis de LaPlace (LaPlace 1902), Bayes' theorem suggests that new evidence should update the estimated probability that a hypothesis is true based on Equation 1, where H represents the hypothesis and E represents the evidence.

Equation 1:

$$P(H|E) = \frac{P(E|H)}{P(E)} \times P(H)$$

P(H) in Equation 1 is the *prior probability*, or the probability that the analyst would have assigned to the hypothesis before the new evidence was available. P(E|H) is the probability of observing the new evidence, given that the hypothesis is true; P(E) is the total overall probability of observing the evidence (regardless of whether the hypothesis is true); and P(H|E) is the *posterior probability*, or the probability that the hypothesis is true, given the evidence that was observed. Thus, when new evidence becomes available, the analyst's prior beliefs should be updated according to the ratio of P(E|H) to P(E).

The degree to which Bayes' Theorem would suggest that a forecaster should update her beliefs in response to new evidence depends on the strength of her prior beliefs. For example, as P(H) approaches either zero or one (i.e. the forecaster already believes that the hypothesis is either definitely true or definitely false), the ratio of P(E|H) to P(E) also approaches one, so that no amount of evidence can have much effect on the posterior probability. This helps to explain Tetlock's (2009) finding that forecasters with a particular cognitive style —characterized by a strong need for closure and a low tolerance for ambiguity— were more likely to assign very high probabilities to the events they predicted and were unlikely to change the beliefs on which they based their predictions, even after their predictions proved to be false. As a result, forecasters who expressed the greatest certainty about their forecasts generally proved to produce the least accurate forecasts. In other words, "There is often a curiously inverse relationship between how well forecasters thought they were doing and how well they did" (Tetlock, 2009, 106-107).

In fact there is a real danger that forecasters will not only fail to update their beliefs in response to new information, but that information that conflicts with deeply-help beliefs may even serve to strengthen those beliefs. This response to contradictory information has been demonstrated in two studies, one on beliefs about the dangers of vaccines (Nyhan et al. 2014) and another on political misperceptions (Nyhan and Reifler 2010).

Methodologies and judgments used to produce a forecast can be evaluated independently from forecast accuracy. Ascher (1979) refers to this type of evaluation as the insider's approach because such an evaluation requires and insider's knowledge of all the assumptions, methods, and judgments that were used to create the forecast, as well as sufficient expertise to determine whether these assumptions, methods, and judgments were appropriate, given the information available when the forecast was made and the current state of practice. Murphy (1993) goes even further and argues that no one besides the forecaster herself can ascertain whether a forecast was consistent with the forecaster's own judgment.

Accuracy and bias

In contrast to evaluating methodology and judgment, evaluating forecast accuracy may seem to be a relatively straightforward task of comparing forecast values to observed values. However, a variety of metrics are available to make such comparisons, and the selection of a metric can influence conclusions about the overall accuracy of particular forecaster or forecasting method. Each of the M competitions (Makridakis et al. 1982; Makridakis and Hibon 2000; Makridakis et al. 1993) evaluated forecast accuracy using five different metrics and found that the relative performance of the specific methodologies they evaluated varied depending on the accuracy metric that was used. A substantial body of literature has emerged to compare and propose metrics for evaluating forecast accuracy. Table 3 lists a selection of these metrics.

Metrics for evaluating forecast accuracy may be either dimensioned or dimensionless. Dimensioned measures are expressed in the units of the quantity being forecast; dimensionless measures are expressed as ratios or percentages.

The most basic dimensioned measure for forecast accuracy is error, which can be calculated as the forecast value minus the actual value. Error will be positive when the forecast is higher than the actual value and negative when the forecast is lower than the actual value. By indicating both the magnitude and direction of the difference between the forecast and the actual value, error is useful for measuring both accuracy and bias for a single forecast. However, for a set of forecasts, average error can indicate average bias, but not average accuracy, since large positive errors have the potential to cancel out large negative errors.

Table 3. Selected metrics for evaluating point forecast accuracy

	Formula[1]	Source
Dimensioned		
Average absolute error (AAE)	$\frac{\sum_{i=1}^{n} \lvert f-a \rvert}{n}$	Described in (Ascher 1979)
Root mean squared error (RMSE)	$\sqrt{\frac{\sum_{i=1}^{n}(f_i - a_i)^2}{n}}$	Described in (Ascher 1979)
Dimensionless		
Mean absolute percentage error (MAPE)	$\frac{\sum_{i=1}^{n}\left(\frac{\lvert f-a \rvert}{a}\right)}{n}$	Described in (Armstrong and Collopy 1992; Kitchenham et al. 2001)
Mean magnitude of error relative to the estimate (MMER)	$\frac{\sum_{i=1}^{n}\left(\frac{\lvert f-a \rvert}{f}\right)}{n}$	Proposed in (Kitchenham et al. 2001)
Symmetric mean absolute percentage error (SMAPE)	$\frac{\sum_{i=1}^{n}\left(\frac{\lvert f-a \rvert}{\left(\frac{1}{2}\right)(f+a)}\right)}{n}$	Proposed in (Armstrong 1978)
Mean z or Q	$\frac{\sum_{i=1}^{n}\frac{f_i}{a_i}}{n}$	Described in (Kitchenham et al. 2001) as z and described in (Tofallis 2015) as Q
Mean ln(Q)	$\frac{\sum_{i=1}^{n} ln\left(\frac{f_i}{a_i}\right)}{n}$	Described in (Lo and Gao 1997), named in (Tofallis 2015)
Percent Better	$\frac{\sum_{i=1}^{n}[\lvert f_i^* - a_i \rvert > \lvert f_i - a_i \rvert]}{n}$	Described in (Makridakis et al. 1982)
Theil's U2	$\frac{\sqrt{\frac{\sum_{i=1}^{n}(f_i - a_i)^2}{n}}}{\sqrt{\frac{\sum_{i=1}^{n}(f_i^* - a_i)^2}{n}}}$	Proposed in (Theil 1966)
Geometric mean relative absolute error (GMRAE)	$\prod_{i=1}^{n} \sqrt[n]{\frac{\lvert f_i - a_i \rvert}{\lvert f_i^* - a_i \rvert}}$	Described in (Armstrong and Collopy 1992)

1. f = forecast value,
   a = actual value,
   f^* = a forecast produced by a "naive" model such as random walk or no-change, and
   n = the number of forecasts being evaluated.

Two common dimensioned measures for the accuracy of a set of forecasts are average absolute error (AAE) and root mean squared error (RMSE) (Ascher 1979). AAE for a set of forecasts is calculated as the difference between the forecast value and the actual value (expressed as a positive number regardless of the direction of the error), averaged across all forecasts. Thus, while average error measures bias rather than accuracy, average absolute error measures accuracy rather than bias. RMSE is calculated by summing the squares of the error for each forecast in a set, then taking the square root of the sum. The RMSE for a given set of forecasts will always be at least as large as the AAE, but RMSE imposes a greater penalty for extreme errors (Ascher 1979). Thus, two sets of forecasts might have the same AAE if one set of forecasts is moderately accurate and the other includes several very accurate forecasts and a few very inaccurate ones. However, the latter set will have a higher RMSE.

A feature of dimensioned measures of accuracy is that they preserve the scale of the error, and this may or may not be an advantage. For example, in cost estimating, an error of three percent might be of greater concern if it represents a difference of one million dollars than if it represents a difference of one hundred dollars. However, this may also make it impractical to compare forecast accuracy across projects or phenomena of widely different scales. Moreover, the amount of error that a forecast user will accept may depend on the scale of the project or phenomenon, so that the ratio of the error to the predicted or forecast value is more important than the magnitude of the error. Beginning in the 1980s, econometric forecasting researchers began to shift from a preference for dimensioned to dimensionless measures of forecast accuracy (Armstrong and Collopy 1992).

The most common dimensionless measure for forecast accuracy is mean absolute percentage error (MAPE) (Tofallis 2015; Armstrong and Collopy 1992; Gneiting 2011). MAPE is calculated as the average of the absolute difference between the forecast and the actual value, divided by the actual value (see the equation in Table 3). Mean magnitude of error relative to the estimate (MMER) is a related measure that divides the absolute error by the forecast value rather than by the actual value (see the equation in Table 3). Kitchenham et al. (2001) argue that MMER is a more useful measure for forecast users than MAPE is, since it allows the user to estimate the likely error that she may expect for new forecasts produced by a forecaster or methodology with an established MMER.

A major flaw of both MAPE and MMER is asymmetry: They both impose a greater penalty for errors in one direction than in the other. For each metric, a value of zero indicates perfect accuracy and higher values indicate less accuracy. However, assuming that the quantity being forecast is strictly positive (i.e. there can be no meaningful negative values), a forecaster or methodology that consistently under-predicts the actual values will have a maximum MAPE of one (or 100 percent), but there will no upper limit on its MMER. Conversely, a forecaster or methodology that consistently over-predicts the actual values will have a maximum MMER of one (or 100 percent), with no upper limit on its MAPE. Thus, MAPE imposes a greater penalty for over-prediction than for under-prediction, and MMER imposes a greater penalty for under-prediction than for over-prediction.

A second, but less serious flaw of MAPE and MMER is that both may be undefined in specific cases. MAPE will be undefined if any actual values are zero; MMER will be undefined if any forecast values are zero.

Armstrong (1978) proposed the symmetric mean absolute percentage error (SMAPE) to address both of these flaws of MAPE and MMER. SMAPE is calculated by dividing the absolute error by the average of the forecast and the actual value (see Table 3). As a result of this modification, the maximum SMAPE for both over-and under-prediction is 2.0 or 200 percent, and the penalty for over-prediction is the same as for under-prediction. For example, if a pessimistic forecaster predicts a value of 50, an optimistic forecaster predicts a value of 200, and the actual value is 100, MAPE would suggest that the optimist was more accurate, with a MAPE of 50 percent compared to the pessimist's MAPE of 100 percent. However, MMER would suggest that the pessimist was more accurate, with a MMER of 50 percent compared to the optimist's MMER of 100 percent. SMAPE resolves this discrepancy with a SMAPE for both forecasters of 67 percent (Armstrong and Collopy 1992).

In some cases, a researcher may want an accuracy measure that preserves information about the direction of the error. Kitchenham et al. (2001) propose that this can achieved by using the average of the ratio of the forecast value to the actual value (see Table 3). Kitchenham et al. (2001) and Lokan (2005) refer to this metric as z; Tofallis (2015) refers to the same metric as Q; and Lo and Gao (1997) have used this metric without naming it. Perfect accuracy will yield an average Q of one; under-prediction will yield values less than one, and over-prediction will yield values greater than one. Q is equal to one plus percent error, and it suffers from the same primary flaw as average percent error: when aggregated to an average, it measures bias rather than accuracy: It is possible for a very inaccurate forecaster or method to achieve an average ratio close to one if forecasts are relatively unbiased. The ratio also has the same problems with asymmetry as MAPE and MMER, although Tofallis (2015) and Lo and Gao (1997) recommend correcting for this by taking the log of the ratio (shown as ln(Q) in Table 3).

An analyst may be interested in whether a forecaster or methodology offers more value than a “naïve” forecast, such as one produced by random guessing or simple rules (e.g., a no-change model). If a set of reference forecasts is available, a few metrics are available to a compare a set of forecasts to the reference forecasts. The simplest of these is Percent Better (see Table 3), which is the percentage of forecasts with a smaller absolute error than the absolute error of the reference forecast. A disadvantage of Percent Better is that it does not account for how much improvement a specific forecast offers relative to the reference forecast (although this may also be an advantage of Percent Better in the sense that it is insensitive to outliers). Theil’s U2 statistic (Theil 1966) and the geometric mean of the relative absolute error (GMRAE) both incorporate information about the magnitude of error. Theil’s U2 is the ratio of the RMSE for a set of forecasts to the RMSE for a set of reference forecasts (see Table 3). GMRAE is the geometric mean of the ratios of the forecasts absolute errors to the reference forecasts absolute errors (see Table 3).

All of the metrics in Table 3 are useful for evaluating point forecasts. However, they cannot be directly applied to forecasts that express uncertainty by giving ranges of possible values or by assigning probabilities to particular events. However, as Murphy (1979) argues, forecasts should be accurate not only in terms of forecast magnitude, but also in terms of the degree of uncertainty associated with the forecast. Since there is never perfect certainty about the future, this notion suggests that forecasts of continuous variables should always be expressed as ranges of possible or likely values.

To assess the accuracy of forecasts that are expressed as ranges, Ascher (1979) takes the midpoint of the range and applies one of the formulae from Table 3, arguing that forecast users are most likely to interpret the midpoint as the most likely value within the range. However, this approach assigns less accuracy to a forecast when the actual value is just within a wide forecast range (in which case the forecast was correct), than when the actual value is just outside a narrow forecast range (in which case, the forecast was incorrect).

Murphy and Winkler (1977) suggest reliability as an indication of the quality of probabilistic forecasts that consist of a probability assigned to a dichotomous event (for example the probability of rain or the probability that the temperature will fall within a particular range). Reliability is the degree to which the forecast probabilities correspond with observed relative frequencies. The most common application of reliability is in weather forecasting. If, when a forecaster predicts a 20 percent chance of rain, it rains 20 percent of the time, the forecaster may be said to be perfectly reliable. The concept of reliability can also be applied to any forecast expressed as a confidence interval. For example, 90-percent confidence interval forecasts are perfectly reliable if observed values fall within the confidence interval 90 percent of the time. Reliability is a useful measure for forecasts of events that occur frequently enough that a meaningful sample of comparison observations can be measured (hence its common application in weather forecasting). However, as Tetlock (2009) notes, probabilistic forecasts of rare events, or events that are conditional on rare circumstances, are much more difficult to evaluate.

Another obstacle to evaluating forecast accuracy arises when decisions made as a result of the forecast have an influence on observed values. This may be a result of self-defeating or self-fulfilling forecasts, as described by Ascher (1979). When performance forecasts are used as a basis for project selection, this also introduces selection bias into estimates of forecast accuracy. Eliasson and Fosgerau (2013) demonstrate that when forecasts of project performance cannot achieve perfect accuracy, and projects are selected based on those forecasts so that accuracy is only observable if the forecast is above a particular threshold, observable errors will always be optimistically biased, even if the errors of the full set of forecasts would have been unbiased had all projects been completed. This effect is illustrated in Figure 1, which represents forecast and actual values for 16 hypothetical proposed projects. Only those projects for which the forecast lies above the selection threshold are selected for completion and thus have observable errors. Based on the observable errors, there appears to be a substantial optimistic bias in the forecasts. However, if all projects had been completed, the full set of forecasts would have been unbiased.

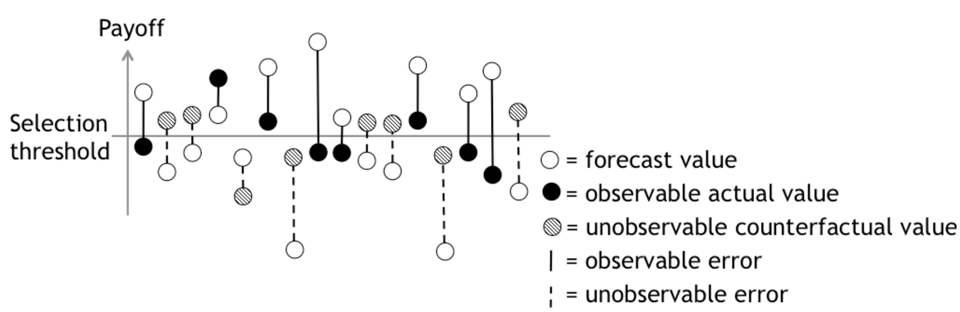

Figure 1. Effect of selection bias on observable forecast accuracy (based on Eliasson and Fosgerau 2013)

Usefulness and value

Murphy (1993) describes the value or usefulness of a forecast as the degree to which the forecast influences forecast users to make better decisions than they would have made without the forecast (or perhaps with a different forecast). Thus, measuring forecast usefulness requires speculation about two counterfactuals that cannot be directly observed:

(1) Would another decision have been made with no (or with a different) forecast?

(2) Would that alternative decision have lead to a better outcome?

Returning to Figure 1, a more accurate forecast would only have led to a different project selection decision for the projects where the line representing the observable or unobservable error crosses the selection threshold. Even when a forecast has substantial error, a more accurate forecast would not have led to a different decision as long as the forecast value and the actual value are both on the same side of the selection threshold.

In the simplified scenario illustrated in Figure 1, the decision is made based exclusively on the forecast. This is not the case for all decisions that might nevertheless be influenced by expert forecasts. Ascher (1979) finds that expert forecasters in fields related to policy and planning tend to be skeptical about the influence of their forecasts on decisions by forecast users, although forecast users themselves "seemed quite convinced of the utility and influence of experts' forecasts" (Ascher 1979, 16). Ascher (1979) attributes this difference, not to differing perceptions of how influential forecasts actually are in decision-making, but rather to differing perceptions of how influential forecasts ought to be in decision-making:

> [T]he forecaster's perspective focuses on securely implanting the forecast in the decision-making routine, which is aided (but not guaranteed) by making consideration of expert forecasts a necessary step in the policy-making process. In contrast, the policy-maker's perspective calls for forecasts (and technical expertise in general) to be useful in his deliberation, but without reducing his flexibility in the policy choice. (Ascher 1979, 17)

When a forecaster concludes that a forecast clearly indicates an optimal decision, but does not trust the forecast user to reach the same conclusion, she may justify providing a less accurate forecast on the grounds that it will lead to a better decision. Tetlock (2009) found that this was a common argument by experts who tended to overestimate the risk of worst-case scenarios: that it was better for decision-makers to over-prepare for the worst than to be surprised by an unlikely, but possible, disaster. On the opposite side of the same coin, Hirschman (1967) has argued that, although project planners often underestimate the costs of proposed projects, they also underestimate the projects' benefits (relative to doing nothing) and the resources that will be available to meet unanticipated costs. Thus, argues Hirschman (1967), it is fortunate when the true costs of a worthwhile project are underestimated, since advance knowledge of the true costs would have discouraged decision-makers from pursuing the project at all. Flyvbjerg (2016) refutes the claim that underestimated costs are accompanied by underestimated benefits. In fact, Flyvbjerg (2016) finds the opposite to be true: when project costs are underestimated, project benefits tend to be overestimated. Thus, he rejects the claim that systematically biased forecasts are desirable to encourage optimal decision-making.

A further danger of consistent bias is its potential to undermine credibility in the long term, as in the story of "The Boy Who Cried Wolf," diminishing the usefulness of the forecast. In response to the perception of persistent bias, decision-makers may seek to counter-balance with additional forecasts from sources perceived to have the opposite bias. De Bruijn and Leijten (2007) describe how such circumstances lead to "contested information" as those for and against a particular decision each produce forecasts that support their respective positions. They argue that the resulting proliferation of conflicting information makes informed decision-making impossible, and undermines the usefulness of all forecasts. As a remedy, De Bruijn and Leijten (2007) suggest that this proliferation may be avoided through the creation of "negotiated knowledge" if, prior to analysis, stakeholders agree on a consistent set of assumptions and methods that will be used to produce forecasts and the ways in which the forecast will inform decision making.

Murphy (1993) argues that, when a forecaster has imperfect information about how the forecast will be used, the usefulness of a forecast is most likely maximized when its accuracy is maximized.

Beyond the aforementioned difficulty in determining whether a different forecast would have led to a different decision, it is likewise difficult to judge whether another decision would have led to a better outcome, a problem that Tetlock (2009) refers to with regard to evaluating past political decisions more generally. For example, if a choice between two competing alternative projects is made based on forecasts of each project's performance, the selected project may underperform so that its actual performance is worse than the forecast performance of the unselected alternative. In that case, it is clear that a more accurate forecast for the selected project would have led to the selection of the other alternative. However, we cannot know that this would have been a better alternative without also knowing how accurate the forecasts for the unselected alternative would have proven to be. If the forecasts for both alternatives had errors of a similar magnitude and in the same direction, the selected alternative would still be the best decision.

Cost estimates for infrastructure projects

In addition to the general characteristics of forecasts discussed in the previous section, a few considerations apply specifically to cost estimates prepared for public works infrastructure projects.

Love et al. (2013) differentiate between two distinct components of the difference between the cost estimate that was made at the point of decision to pursue a project and the total capital expenditures that have been made by the time a project is completed. First, there is cost escalation, which is the difference between the initial cost estimate and the value of awarded construction and procurement contracts. Second, there are cost overruns, defined as the difference between the original contract values and the final contract awards that have been paid by the time the project has been completed. It is noteworthy that the terms “cost escalation” and “cost overruns” refer to characteristics of observed costs rather than to characteristics of cost estimates. This terminology emphasizes a failure to stay within the budget, rather than a failure to accurately predict costs, and may thus shift blame from forecasters and estimators to designers and project managers.

Trost and Oberlender (2003) find that five factors affect the accuracy of early cost estimates: basic process design, team experience and cost information, time allowed to prepare the estimate, site requirements, and bidding and labor climate.

The competitiveness of the contract bidding environment, as measured both by the number of bidders (Rowland 1981), the difference between the engineering cost estimate and the low bid (Jahren and Ashe 1990), and the difference between the high bid and the low bid (Hinze, Selstead, and Mahoney 1992), has been demonstrated to be associated with the likelihood of cost overruns since bidders are motivated to build fewer contingencies into their bid amounts in order to secure the project, and must make up for this through change orders when unexpected circumstances arise. Greater project complexity (Rowland 1981; Hinze, Selstead, and Mahoney 1992) and size (Jahren and Ashe 1990) also are associated with a higher probability of cost overruns.

Liu and Zhu (2007) propose five different project phases during which cost estimates may be prepared: the conceptual stage, the design stage, the tender stage, the preconstruction stage, and the build stage. These stages apply to a traditional design-bid-build project delivery method, rather than a design-build method. The level of accuracy that is achievable and desirable increases with each successive stage.

During the conceptual stage, minimal effort is put into cost estimation, since the inputs to the estimates are highly uncertain and the objective of the estimate is simply to determine if the proposed project is adequately feasible to warrant further study (Liu and Zhu 2007). The accuracy of cost estimates prepared during the conceptual stage ranges from 30 to 50 percent (Liu and Zhu 2007; Ashworth and Skitmore 2005). Liu and Zhu (2007) suggest that the most effective way to improve the accuracy of cost estimates prepared during the conceptual phase is to choose experienced, skilled people to prepare cost estimates (or improve the knowledge and skills of available cost estimators), improve the quality of input data, and integrating early design tasks with early cost estimation tasks.

The experience of the cost estimator continues to be of primary importance for the accuracy of cost estimates prepared during the design stage (Liu and Zhu 2007). The objective of design-stage cost estimation is to evaluate the feasibility of proposed designs within a budget constraint (Liu and Zhu 2007). The accuracy of estimates prepared at this stage are generally around 20 percent (Liu and Zhu 2007; Morrison 1984).

During the tender stage, the objective of cost estimation is to prepare to make a competitive (and realistic) bid on a project (Liu and Zhu 2007). Thus, estimate accuracy becomes more important, and improves to about 10 percent (Flanagan and Norman 1983; Liu and Zhu 2007). Furthermore, since most of the design has usually been completed by this stage, the effect of subjective judgment (and hence estimator skill and experience) on cost estimate accuracy diminishes (Liu and Zhu 2007). Cost estimate accuracy at this stage can be improved by setting standards and expectations for the degree of accuracy that must be achieved and by establishing clearly-defined procedures for preparing cost estimates (Liu and Zhu 2007).

The objective of cost estimates prepared during the preconstruction phase is to make accurate budget predictions (Liu and Zhu 2007). By this point, project scope is very well defined and accurate cost estimates can most effectively be improved by setting standards and expectations for the required accuracy, which is typically around five percent (Ferry, Brandon, and Kirkham 2014; Liu and Zhu 2007).

During the build phase, the emphasis generally shifts from predicting costs to containing costs. Thus, deviations from cost estimates prepared during the preconstruction phase would be categorized as cost overruns rather than cost escalation. In general, cost overruns have been observed to be relatively modest compared to cost escalation that occurs prior to a contract award (Love et al. 2015; Hinze, Selstead, and Mahoney 1992; Jahren and Ashe 1990).

Siemiatycki (2009) describes how two groups —academics and government auditors— who study inaccuracy in cost estimates for transportation projects emphasize different explanations for cost overruns. Academics, particularly those from planning and other social science disciplines, tend to emphasize strategic misrepresentation and other political and psychological explanations. Auditors, and also academics in the engineering and construction management disciplines, are more likely to emphasize technical explanations for cost overruns. This difference illustrates the importance of taking an interdisciplinary approach to understanding errors in cost estimation, since both technical factors and political or psychological factors likely play a role in cost estimate errors.

While Kahneman (Kahneman and Tversky 1979; Kahneman 2013) and Flyvbjerg (Flyvbjerg, Holm, and Buhl 2002; Flyvbjerg 2006) both recommend reference class forecasting as a response to optimistic bias in cost estimation, each ascribes different explanations to it. Kahneman describes the appraisal optimism that comes from taking the inside view as a common and innocent mistake on the part of cost estimators (Kahneman 2013). Flyvbjerg and colleagues (Flyvbjerg, Holm, and Buhl 2002) acknowledge that sincere mistakes would be innocent: “Such optimism, and associated cost underestimation, would not be lying, needless to say, because the deception involved is self-deception and therefore not deliberate" (Flyvbjerg et al., 2005, 288). In other words, to quote George Costanza, "It's not a lie if you believe it" (Ackerman 1995). However, they go on to suggest that innocent mistakes cannot be an important source of error in transportation cost estimates since many estimators are experienced and can be expected to have learned from their mistakes and those of others over time.

> Appraisal optimism would be an important and credible explanation of underestimated costs if estimates were produced by inexperienced promoters and forecasters, i.e., persons who were estimating costs for the first or second time and who were thus unknowing about the realities of infrastructure building and were not drawing on the knowledge and skills of more experienced colleagues. Such situations may exist and may explain individual cases of cost underestimation. But given the fact that the human psyche is distinguished by a significant ability to learn from experience, it seems unlikely that promoters and forecasters would continue to make the same mistakes decade after decade instead of learning from their actions. It seems even more unlikely that a whole profession of forecasters and promoters would collectively be subject to such a bias and would not learn over time. Learning would result in the reduction, if not elimination, of appraisal optimism, which would then result in cost estimates becoming more accurate over time. (Flyvbjerg et al., 2002, 288-289)

As discussed in the previous section, the argument that the human psyche's significant ability to learn from experience rules out the possibility for unintentional error on the part of experienced forecasters may present too optimistic a view of human psyche, since prior beliefs and biases have proven to be difficult to change, even in the face of new evidence (Nyhan and Reifler 2010; Nyhan et al. 2014; Tetlock 2009).

Demand forecasts for transportation infrastructure

Many of the social and political explanations for inaccurate cost estimates may be (and have been) applied to inaccurate demand forecasts as well (Flyvbjerg, Skamris Holm, and Buhl 2005). However, evaluating the accuracy of demand forecasts can be more challenging since there can be uncertainty associated with both the forecast values and the actual values. Uncertainty in actual demand values may be a result of measurement error. Uncertainty can also result from temporal variation in demand, since a measurements of actual demand will vary depending on the particular days and times and which demand measurements are made.

Measurement error is likely to play an important role in ridership forecasts. One major problem faced by agencies seeking to forecast future ridership is the difficulty of ascertaining existing ridership. A 2006 survey of transit operators found that a majority of survey respondents do not have adequate ridership data to forecast future ridership, particularly at the sub-route level (e.g. for route segments or individual stops) (D. Boyle 2006). Respondents further reported lower levels of confidence in the reliability of ridership data than in the reliability of other model input data (D. Boyle 2006).

To further complicate matters, transit ridership forecasts are typically made in terms of linked trips, while ridership data are more likely to be collected and reported in terms of unlinked trips (D. Boyle 2006). Attempts to convert between the two introduce an additional source of error and uncertainty. There are both theoretical and practical reasons for selecting one measure over the other. It is usually easier to collect data on unlinked trips, since this can be done through manual or automated counts of boardings and/or alightings (D. Boyle 2006). However, unlinked trips must be estimated using three of the four steps in the traditional four-step model for travel demand forecasting (trip generation, trip distribution, and mode split). Since uncertainty increases with each step of the model (Zhao and Kockelman 2002), estimates of linked trips are likely to be more accurate than estimates of unlinked trips. The reverse is true for the actual measurement of linked and unlinked trips.

The distinction between linked trips and unlinked trips is not trivial, since some transit system changes can have opposite effects on the two measures. For instance, if service is changed to require more transfers (which tends to make riding transit more onerous), this could simultaneously reduce the number of door-to-door unlinked trips by travellers while increasing the number of unlinked trips (boardings) by those fewer travellers (Taylor et al. 2009).

Even if transportation system utilization could be measured with perfect accuracy, measurements would vary somewhat from day-to-day. It is common to express transportation system demand in terms of average annual weekday values, but these values are rarely accompanied by measures of spread, such as 95-percent confidence intervals or standard deviations, which would facilitate an estimation of how often and by how much demand might exceed these averages. Moreover, annual averages can vary slightly depending on the date chosen to begin the year in which they are measured.

A few researchers have compared the accuracy of demand forecasts among different types of transportation infrastructure projects. In a study comparing demand forecasts for road and rail transportation projects in Europe, Næss, Flyvbjerg, and Buhl (2006) found that demand forecasts for road projects tend to be more accurate than rail project forecasts. This difference could be a function of the low mode shares that are typical on transit, including rail transit. When rail mode shares are much lower than road mode shares, a small error in the overall share of travelers that use rail will result in larger percent errors in demand for rail travel than in demand for road travel.

Differing measures of success for different types of projects may also play a role in creating different incentives for biased forecasts. For example, both rail projects and road projects may be justified by a goal of reducing roadway congestion, either by increasing roadway capacity (through a road project) or by shifting a roadway's vehicle trips to rail trips (through a rail project). For rail projects, success at achieving congestion relief seems most likely if ridership forecasts are high. For road projects, success at achieving congestion relief seems most likely if traffic volume forecasts are low. Indeed, Parthasarathi and Levinson (2010) have found that, in a sample of traffic forecasts for roadway projects in Minnesota, there has been a tendency to underestimate traffic volumes. Toll roads create a different set of incentives than non-tolled facilities, since project promoters must show that the project will generate enough toll revenue for the project to be financially feasible. In a study of tolled and non-tolled roadway projects in Norway, Welde and Odeck (2011) found that forecasts for toll roads tended to be more accurate than those for non-tolled roads, and attributed this difference to a failure of forecasters to account for induced traffic on non-tolled roads and a greater degree of scrutiny given to toll road demand forecasts. Bain (2009) reports a similar finding in a meta-analysis of global toll road projects.

Cost and ridership forecasts for transit infrastructure in the United States

As described in Chapter 1, Federal grant programs providing capital funding for fixed-guideway public transit projects began in the 1960s and 1970s. Beginning in the 1980s, many observers began to question this central federal role in transit funding. Some critics supported the Reagan Administration's New Federalism initiatives and argued that transit funding was one of many Federal programs that would be better funded and managed at the state and local levels (Davis 2015). Others further argued that urban rail transit was generally a bad investment —or a risky investment at least— and that the Federal government encouraged inappropriate risk-taking by placing the investment decision in the hands of local decision-makers while assuming a substantial portion of the risk at the Federal level (Gomez-Ibanez 1985a). It was in this political environment that the earliest studies of the accuracy of forecasts prepared for federally funded transit infrastructure projects were published.

Gomez-Ibanez (1985b) examines the performance of the first three modern (post-1950) light-rail systems to be constructed in North America, including two in Canada (Edmonton and Calgary) and one in the United States (San Diego), all of which were completed between 1978 and 1981. Rather than comparing project costs and performance to specific forecasts prepared for those projects, Gomez-Ibanez (1985b) compares performance to general claims light rail advocates had made about the benefits of light rail: that, for modestly higher capital costs than bus service, light rail can attract more transit passengers and serve those passengers at a lower operating cost per passenger mile. He finds neither increases in ridership nor reductions in operating costs, relative to his estimates of what the bus lines the new systems replaced would have achieved. Gomez-Ibanez's (1985b) argument thus rests on assumptions about counterfactuals. As summarized earlier in this chapter Tetlock (2009) and Eliasson and Fosgerau (2013) each discuss how the inability to observe counterfactuals can make it difficult to evaluate past decisions. However, since all three cities hosting the light rail projects Gomez-Ibanez (1985b) evaluates had existing bus service that continued after limited replacement by light rail service, his estimates of how the discontinued routes would have performed are likely to be reliable.

Kain (1990) presents a case study of ridership forecasts prepared for the first urban rail system in Dallas. Since the urban rail system had not yet opened for service and the forecasts were for the years 2000 and 2010, Kain's (1990) study could not take Ascher's (1979) "outsider's approach" to forecast evaluation by evaluating measure forecast accuracy. Instead, Kain (1990) takes Ascher's (1979) "insider's approach," focusing on the forecast methodology, with an emphasis on assumptions for inputs such as highway congestion, central business district employment levels, and parking costs. Kain (1990) finds that these assumptions were largely incorrect and points to evidence that the selection of input assumptions was politically motivated to produce inflated ridership forecasts.

In his criticism of the ridership forecasts for the first urban rail lines in Los Angeles, Richmond (2005) likewise takes Ascher's (1979) "insider's approach" to forecast evaluation, and goes on to claim that empirical tests of forecast accuracy are irrelevant since traditional methods of travel demand modeling are so inadequate that the accuracy of the forecasts they produce can only be coincidental.

Kain (1990) stops short of presenting an explanation for Dallas officials' consistent efforts to use intentionally misleading analysis to justify a project that was not economically viable:

> "While some advocates were clearly acting out of perceived self-interest, the unswerving and blind commitment of many others to rail is difficult to explain in these terms. I leave it to others, more skilled in bureaucratic and political analysis of psychology to provide an explanation" (193).

Richmond (2005) attempts to do just that in his exploration of why the Blue Line of the light rail system in Los Angeles was politically popular in spite of being (In Richmond's view) economically unjustified. He argues that rail transportation has enormous appeal as a symbol of success, growth, renewal, and even sexual power, and that this symbol is so powerful that it defies refutation by evidence suggesting that rail will not benefit a community. According to Richmond's (2005) study, local decision-makers in Los Angeles did not use ridership forecasts as a basis for evaluating potential urban rail projects, but rather as a game or ritual that must be completed in order to proceed with their pre-determined action.

Wachs (1990) likewise describes the influence of politics on forecasting for transit projects. After interviews with "public officials, consultants, and planners" in which many shared stories of being pressured to revise forecasts in support of politically popular projects, Wachs (1990) concludes,

> "I am absolutely convinced that the cost overruns and patronage overestimates were not the result of technical errors, honest mistakes, or inadequate methods. …The forecasts had to be "cooked" in order to produce numbers which were dramatic enough to gain federal support for the projects whether or not they could be fully justified on technical grounds" (144).

The first study to empirically evaluate forecast accuracy for several urban rail transit projects in the United States was completed for the Urban Mass Transit Administration (UMTA) in 1989 and examined cost and ridership forecasts for four light-rail, four heavy-rail, and two downtown people mover projects that had been completed with UMTA funds between 1983 and 1987 (Pickrell 1989). Pickrell (1989) argues:

> "If the divergence between a project's forecast and actual cost-effectiveness in attracting new transit passengers exceeds the margin by which the chosen alternative was preferred to others that were rejected, the planning process may not have led to selection of the most desirable project" (vi)

This argument rests on two assumptions. First, it assumes that the forecasts for the alternatives not selected offer a good approximation of how they would have performed, had they been selected. As discussed by Tetlock (2009) and Eliasson and Fosgerau (2013), the accuracy of a forecast for an alternative that is not selected is unobservable. However, in the 1980s, there was much less history of forecasting cost and demand for urban rail service than for its most likely alternatives, such as bus service and highway capacity. As discussed in the prior section, previous research has found demand forecasts to be more accurate for highway projects than for rail projects (Næss, Flyvbjerg, and Buhl 2006). Because there is now so much experience with the outcomes of bus projects and highway capacity expansions, it may be that the forecasts for these alternatives would today be more accurate than are forecasts for the rail projects eventually built. Despite the far greater frequency of bus vis-à-vis rail service, there has been less research on the accuracy of bus ridership forecasts than on rail ridership forecasts.

Second, Pickrell's argument also assumes that forecasts have an important influence on decision-makers' selection of a preferred alternative. This assumption may seem to conflict to some degree with claims by other researchers suggesting that the early selection of an alternative (based on considerations other than performance forecasts) influences forecasts to a greater degree than forecasts influence the selection of an alternative (Kain 1990; Richmond 2005; Wachs 1990). However, the two claims are not mutually exclusive, since "decision-makers" are not a monolithic group: project promoters may successfully insert optimism into forecasts in order to win over more skeptical stakeholders.

In response to Pickrell's (1989) finding that cost estimates for many of the rail projects that had been completed had proven to be overly optimistic, transportation consultant Julie Hoover argued that more accurate forecasts would not have changed the decision by local decision-makers, an interesting perspective on the apparent irrelevance of forecasts from a person who produced forecasts for many years. She questioned:

> Does anyone in this audience seriously believe Portland officials would have voted differently if they thought the cost of the [first light rail] Project was $210 million rather than the $172 million originally projected? … Who here thinks that [Atlanta] Mayor [Andrew] Young might be alarmed by the Pickrell report, and wish that Atlanta had opted for an all-bus system if he had only known from the beginning that the costs of rail would be higher? (quoted in Demery Jr. and Setty 2007, para. 70-72)

Pickrell (1989) finds that forecasts for ridership on the new rail systems exceeded actual ridership by 28 to 85 percent, with an average of 65 percent; forecasts of capital costs exceeded actual costs by 17 to 156 percent, with an average of 77 percent.

In order to estimate the proportion of the error in ridership forecasts that can be attributed to errors in forecasting the models' inputs, Pickrell (1989) multiplies the percent error for each of eight input variables (rail headway, rail operating speed, rail fare, feeder bus headway, auto operating cost, parking cost, population, and employment) by their respective estimated elasticities to determine the percent error that would be expected based only on errors in those input variables. Based on this analysis, Pickrell (1989) concludes that errors in input variables accounted for less than half of the error in the total ridership forecasts, and that the remaining error must then be explained by "less obvious sources, including the structure of the … models …, the way in which they were applied, or the misinterpretation of their numerical outputs during the planning process" (Pickrell 1989, 29). One way that model structure can contribute to forecast error without necessarily being technically incorrect is by magnifying input errors in each step of a complex model (Zhao and Kockelman 2002)

For capital cost, Pickrell (1989) likewise finds that unanticipated inflation and changes in project scope and schedule can explain no more than one third of cost overruns for any of the ten projects studied.

In an article discussing the accuracy of cost and ridership forecasts for a subset of the projects discussed in the 1989 report (Pickrell 1989), Pickrell (1992) suggests:

> "The most effective way to induce planners and decision-makers to choose projects on the basis of more accurate ridership and cost projections would be to transfer the financial risk of forecasting error from the federal treasury to local government" (Pickrell 1992, 170).

This suggestion implies that the explanation for inaccurate cost and ridership forecasts lies not in the *inability* to produce accurate forecasts, but rather from the *motivation* to produce inaccurate ones. This assessment is consistent with the arguments of several other scholars (Gomez-Ibanez 1985b; Wachs 1990; Richmond 2005; Kain 1990).

Pickrell (1992) also suggests that cost and ridership forecasts be expressed as confidence intervals rather than as point estimates. This is consistent with Murphy's argument that a good forecast must represent the forecaster's best judgment, not only with regard to the prediction itself, but also with regard to the uncertainty associated with that prediction.

A 2003 study (Spielberg et al. 2003) and a 2008 study (Lewis-Workman et al. 2008) by the FTA follow up on Pickrell's (1989) study by comparing his findings with observations of projects that had been completed between 1990 and 2002 and between 2002 and 2006. Both studies found that the accuracy of cost estimates and ridership forecasts completed after Pickrell's (1989) study was better than that of the forecasts completed before the 1989 study. The authors of the 2003 study (Spielberg et al. 2003) found that unanticipated inflation was a primary source of error in cost estimation and that, when cost estimates were adjusted to reflect inflation, estimates were generally within 20 percent of actual costs. They also suggest that some of the improvement in cost estimate accuracy can be explained by reduced delays in project development and construction — and the resulting shorter time period between cost estimation and project opening. With regards to the accuracy of ridership forecasts, Spielberg et al. (2003) suggest four reasons for observed improvements: increases in experience, greater scrutiny, improved forecasting methods, and improvements in computing power (although, as described earlier in this chapter, Alonso (1968) argued that the increased model complexity that relies in increased computing power may serve to exacerbate forecast error). They also find that ridership forecasts for initial lines of new systems have been less accurate than expansions of existing systems and that forecasts for transitways and downtown people movers were particularly inaccurate.

In spite of the gains documented by Spielberg, et al. (2003), Lewis-Workman et al. (2008) did not find that the accuracy of cost estimates or ridership forecasts had continued to improve in the five years following Spielberg et al.'s (2003) study.

Both Spielberg et al. (2003) and Lewis-Workman et al.(2008) find that actual service levels have generally been well below those assumed when generating ridership forecasts, but note that it is not clear whether service was reduced in response to low ridership or ridership was lower than anticipated in response to lower-than-anticipated service.

Of the three studies discussed above (Pickrell 1989; Spielberg et al. 2003; Lewis-Workman et al. 2008), none included tests of the statistical significance of changes in the accuracy of cost estimates and ridership forecasts over time or differences in cost estimate accuracy by project characteristics. Such an analysis may not have been possible since, at the time each study was completed, too few federally funded transit projects had been constructed to allow for such an analysis to be meaningful or informative. One study that did test for improvements in cost estimate accuracy (measured as MMER, as defined in Table 3) using a difference of means test for projects completed before and after the Pickrell study failed to find an improvement that was significant at a 95-percent confidence level (Dantata, Touran, and Schneck 2006). Subsequently, Button (2009) likewise tested for a "Pickrell effect" on the accuracy of cost estimates and ridership forecasts (measured as AAE, as defined in Table 3) using a regression analysis to control for project mode, station density, and whether the project was an expansion of an existing system. Button (2009) found that only the presence of an existing system was a significant predictor of ridership forecast AAE at a 95-percent confidence level, and that project mode was also significant when AAE was log-transformed. None of the variables included in Button's (2009) model predicting the accuracy of cost estimates were significant at a 95-percent confidence level.

Schmitt (2016a) has compiled a database of ridership forecast accuracy for transit projects in the United States that is intended for use by travel demand modelers who wish to apply the principles of reference-class forecasting suggested by Kahneman and Tversky (1979) and Flyvberg (2006). In order to define appropriate reference classes, Schmitt (2016a) tests for differences in ridership forecast accuracy by project transportation mode, phase of project development (Alternatives Analysis, Preliminary Engineering, or Final Design), year of construction (defined as either before or after 2007, when the FTA introduced new analytical tools that increase model scrutiny), and whether the project represents a new system, a new line, or an extension of an existing line. He found statistically significant differences only by project transportation mode (forecasts for light rail projects tended to be more accurate than those for projects of other modes) and by whether a project was constructed before or after 2007 (forecasts for older project tended to be less accurate).

Conclusion

Returning to the broader literature on forecasting, three forecast characteristics are relevant to forecast evaluation: the methodology and judgment used to produce the forecast, the accuracy of the forecast, and the usefulness of the forecast. As described in the earlier discussion of these three characteristics, these three characteristics are closely related. Sound methodologies and judgment are necessary to consistently produce accurate forecasts, and an inaccurate forecast cannot be useful. Thus, while the focus of this dissertation is on forecast accuracy, methodology, judgment, and usefulness ought not be ignored. The following chapter describes the mixed-methods approach I have used to examine the context in which forecasting for transit projects take place, including the ways in which methodologies, factors influencing forecaster judgment, and perceptions about the use and usefulness of forecast have changed over time.

From there, I build on the research literature reviewed here by examining trends in the accuracy of cost estimates and ridership forecasts for a larger set of federally funded transit projects than was available at the time of previous studies. While previous studies evaluated the accuracy of individual New Starts projects and documented patterns in accuracy that became apparent early in the years of the New Starts program, a sufficient sample of projects to allow for a comprehensive analysis of trends in and correlates of accuracy for both cost and ridership forecasts has not been available until quite recently, and the lack of such an analysis represents a gap in the existing literature. Likewise, while previous research has qualitatively evaluated the context in which forecasting takes place through individual case studies (Kain 1990; Richmond 2005), there has been no such qualitative, contextual analysis of the overall history of the New Starts program. This dissertation fills these gaps in the literature, as described in Chapter 4.

Chapter 4. Conceptual Framework and Methodology

Much of the prior research evaluating the accuracy of forecasts for federally funded transit projects in the United States was completed relatively early in the history of the New Starts program. As a result, these early studies do not account for the many ways in which the New Starts program has evolved over the years (indeed, many of these changes were motivated in part by the findings presented in these earlier studies). Accordingly, this chapter describes my hypotheses for how changes in the New Starts program relate to changes in forecast accuracy and presents the methodology I use to test these hypotheses.

Conceptual Framework

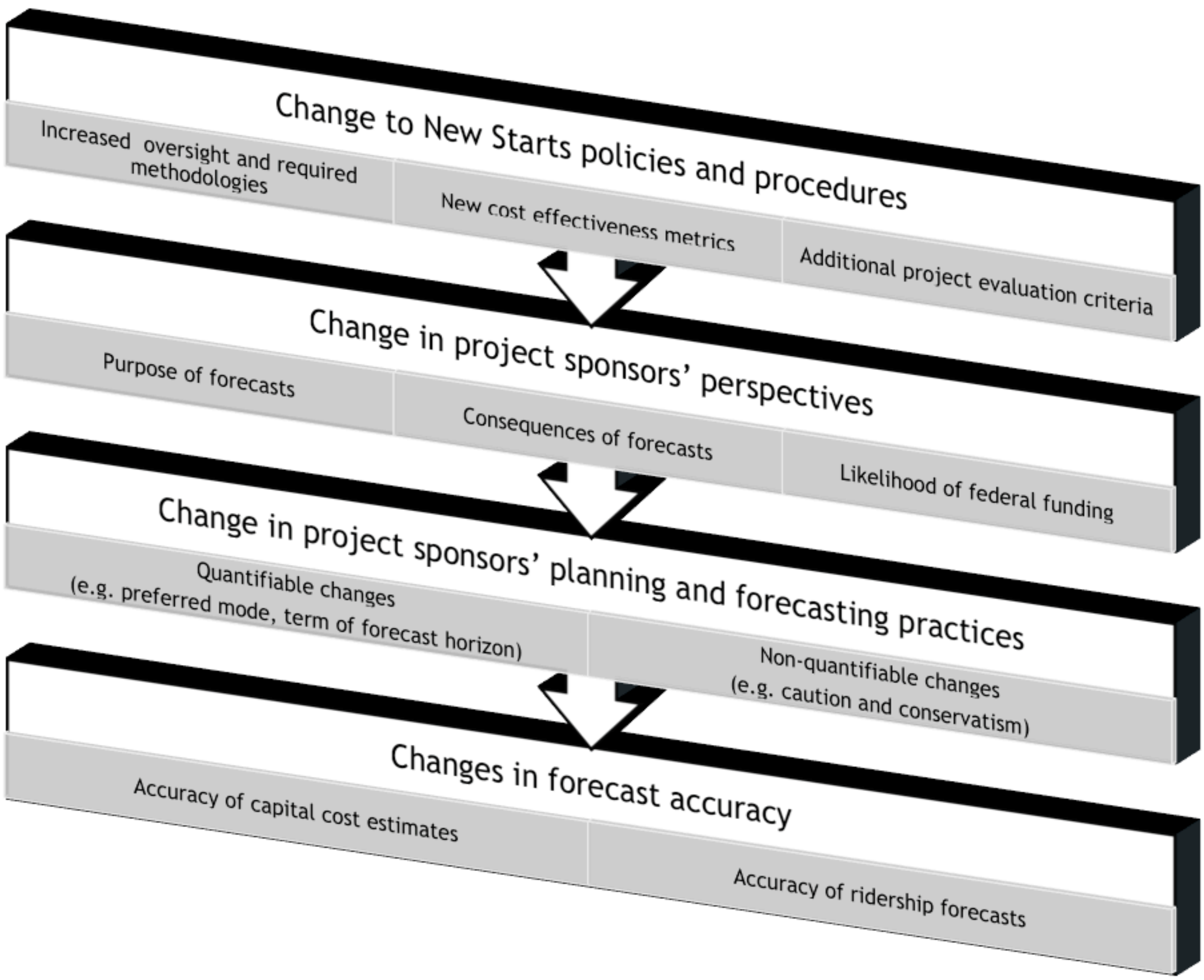

Figure 2: Hypothesized relationship between New Starts policies and forecast accuracy

I hypothesize that changes in cost and ridership forecast accuracy for rapid transit projects can be understood in the context of changes to the New Starts program, as illustrated in Figure 2.

As the FTA (or its predecessor UMTA) has changed the cost-effectiveness metric used to evaluate proposed projects, added additional project evaluation criteria, increased oversight throughout the project development process, and made other changes to the New Starts program, these changes may affect project sponsors' perceptions of the New Starts process. For example project sponsors might change their perceptions of the purpose of the forecasts they prepare (both in terms of how they are used by FTA and how they could be used locally), the consequences of forecast optimism, and the overall likelihood that a particular project will be awarded federal funds. These changing perspectives on the part of project sponsors might affect their actions in both quantifiable and non-quantifiable ways.

There are several non-quantifiable ways in which transit planning and forecasting practice could change in response to changing perceptions on the part of project sponsors. For example, they might increase or decrease their attention to detail and their consideration of uncertainty while preparing forecasts; they might change the level of sophistication or conservatism in their models and underlying assumptions; or they might change the degree to which they consider local political factors in selecting a project. Although quantitative metrics could conceivably be developed to describe some of these possible responses to the changing perceptions of project sponsors, such metrics would likely not be comparable across project sponsors.

Other changes in transit planning and forecasting practice that might result from changing perspectives by project sponsors are both quantifiable and directly measurable. For example, one can observe whether project sponsors complete projects more quickly, prepare forecasts in reference to an earlier horizon year, or pursue smaller-scale projects rather than larger-scale projects.

Both quantifiable and non-quantifiable changes in transit planning and forecasting practice could influence the accuracy of cost and ridership forecasts. This might occur through one or both of two distinct mechanisms. First, accuracy could improve through project design and selection. Cost and ridership may be easier to predict for some types of projects than others. If project sponsors respond to changes in the New Starts program by favoring projects for which it is either more or less difficult to produce reliably accurate forecasts, one would observe a corresponding change in forecast accuracy. Second, accuracy could change due to changes in forecasting practice.

Methodology

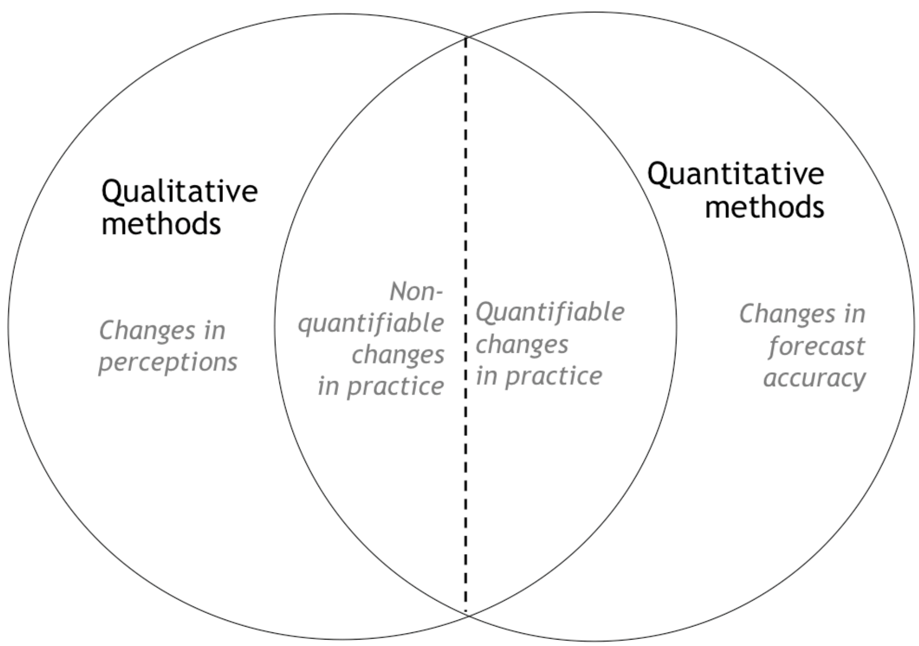

Figure 3: Correspondence between hypotheses and research methods.

Changes in perspective can best be observed through qualitative methods, and many changes in practice are more reliably observed and described qualitatively than quantitatively. In contrast, some changes in practice can be measured and described quantitatively, and changes in forecast accuracy over time are best described quantitatively. Thus, a mixed-methods approach is required to test the hypothesized changes described in the previous section, as illustrated in Figure 3.

To identify changes in perspective and qualitative changes in practice, I conducted a series of interviews with transit professionals. I measured quantifiable changes in practice and accuracy through an analysis of New Starts projects that opened between 1983 and 2011.

Interview methodology

The purpose of the interviews was to identify the ways in which (1) perspectives about the New Starts program have changed over time and (2) transit planning and forecasting practice have changed in response to these changes in perspective.

I interviewed 13 transit professionals, including current and past staff from six transit agencies, three consulting firms, and FTA (or its predecessor, the Urban Mass Transit Administration). All interviewees had experience in preparing, supervising the preparation of, or evaluating ridership forecasts or cost estimates for New Starts projects. Collectively, the interviewees have had a total of over 300 person-years of experience in the transit industry.

I employed a snowball sampling method to recruit interviewees. I recruited three interviewees by reaching out to staff members at transit agencies or consulting firms that had completed forecasts for multiple New Starts projects in the past fifteen years. The other ten individuals I interviewed were referrals from other interviewees or from people who declined to be interviewed but suggested a colleague who might be in a better position to offer relevant insights. Of the nineteen people whom I invited to be interviewed, three declined (two citing a lack of relevant expertise, and one citing a lack of available time), and three failed to respond to the invitation.

In general, a potential weakness of a snowball sampling method is that it can limit the sample to a closed social network of people while excluding those with relevant perspectives who may not have contacts with that network. In this instance, that risk is minor, since the transit planning and forecasting communities are relatively small and offer many opportunities for professional interaction and exchange of ideas. In order to mitigate the risk of limiting the sample to a small, closed network, I took care to ensure that the sample represented a variety of types of organizations (small and large consulting firms, local transit agencies, and a federal agency), a variety of professions (engineers, planners, and managers), a variety of ages (those who had entered the profession within the past decade and those who had retired after long careers), and a variety of geographic regions.

Each interview lasted between thirty and sixty minutes. I completed all interviews by telephone or via web-based video chat, and I provided interviewees with a list of interview questions (included in Appendix A) in advance of their interviews. Interviews were semi-structured in the sense that I used the list of interview questions included in Appendix A as a guide rather than as a script. In many cases, I skipped over questions that were not relevant to an interviewee's background in favor of more detailed follow-up questions on topics about which that the interviewee had more to say. With the permission of each interviewee, all interviews were audio-recorded and transcribed. Upon completion of all 13 interviews, I reviewed all transcripts to identify specific insights interviewees had brought up, grouped the insights into common themes and topics, and selected quotations from the interview transcripts that either illustrated a point that had been made by multiple interviewees or that offered a unique perspective as a counterpoint to the general consensus. Interviewees were given the opportunity to review any direct quotations that appear in this dissertation. In some cases, interviewees chose to edit their quotations for clarity, while maintaining the substance of their statements from the original interviews. In other cases, the interviewee requested that a particular quotation not be used, either because he or she believed it was specific enough to identify him or her (or his or her employer or client), or because it expressed an opinion that he or she did not wish to be published. In all cases, when an interviewee declined to have a direct quotation included, the quotation in question was one of multiple quotations expressing a similar perspective, so I replaced the excluded quotations with alternative quotations that had been approved by the interviewees who had given them.

Quantitative analysis data and methods

Table 4: Relationships tested through quantitative analysis

	Time of first forecast		Accuracy		
			Cost estimate		Ridership forecast
	Year	Governing legislation	Initial	Final	
Time variables					
Year	NA	NA	a, d	a, d	a, d
Governing legislation	NA	NA	b, d	b, d	b, d
Changes in planning and forecasting practice					
Elapsed time					
Project development	a	b	a, d	a, d	a, d
Project construction	a	b	a, d	a, d	a, d
Ridership forecast to horizon year	a	b	e	e	a, d
Financial characteristics					
Total project cost	a	b	a, d	a, d	a, d
Federal funding share	a	b	a	a	a, d
Physical characteristics					
Project length	a	b	a, d	a, d	a, d
Project mode	b	c	b, d	b, d	b, d
Agency experience					
Project share of forecast-year system miles	a	b	a, d	a, d	a, d
Project sequence	b	c	b, d	b, d	b, d
External control variables					
Change in fuel price	e	e	e	e	d
Change in unemployment	e	e	e	e	d
Change in population	e	e	e	e	d
Years from horizon to observation year	e	e	e	e	d
Forecast-year unemployment	e	e	d	d	d
Forecast-year population	e	e	d	d	d

Method of testing relationship
a. Pearson's correlation
b. Tukey's Honestly Significant Difference (HSD)
c. Cramer's V and Pearson's chi-squared
d. Linear regression with controls for all variables listed in left column
e. Relationship not tested

The purpose of the quantitative analysis was to quantify the degree to which (1) transit planning and forecasting practice have changed over the course of the New Starts program and (2) these quantifiable changes in transit planning and forecasting practice can explain changes in the accuracy of capital cost and ridership forecasts.

The sample of projects included in the quantitative analysis comprises all federally funded fixed-guideway transit projects completed between 1983 and 2011 for which data are available on the accuracy of cost and/or ridership forecasts prepared early in the project development process: a total of 67 projects. Unless otherwise noted, project data were compiled from Pickrell (1989), two predicted-versus-actual studies published by the FTA (Spielberg et al. 2003; Lewis-Workman et al. 2008), and summaries of completed Before and After Studies published by FTA (2011b, [b] 2012, [b] 2013, [b] 2015, 2016).

Table 4 lists the variables that are included in the quantitative analysis, the relationships I tested between them, and the methods that I used to test each relationship. In the following subsections, I describe the methods used to test each relationship, followed by a description of each variable in Table 4. Appendix B includes the value of each variable for each project and the source for each value.

Methods for testing relationships

To determine the degree to which planning and forecasting practice for New Starts projects have changed over time, I tested simple relationships between each time variable and each variable describing planning and forecasting practice, as shown in Table 4. The methods used to test these relationships varied depending on whether they related to categorical or continuous variables. To test for a simple relationship between two continuous variables, I used Pearson's correlation, which indicates the strength, direction, and significance of the relationship. To test for a simple relationship between a categorical variable and a continuous variable, I used Tukey's Honestly Significant Difference (HSD) test, which indicates strength, direction, and significance in means among observations in each of multiple categories. To test for a simple relationship between two categorical variables, I used Cramer's V to determine the strength of the association together with the p-value for Pearson's chi-squared to determine the significance of the association.

To determine the degree to which changes in planning and forecasting practice may have influenced the accuracy of cost and ridership forecasts, I estimated series of linear regression models for each of the three accuracy variables shown in Table 4. For each accuracy variable, I estimate a constant model with no independent variables and a control model including only the external control variables shown in Table 4. The change in model fit between the control model and a full model including all of the variables listed in Table 4 indicates whether (and the degree to which) the full set of project characteristics help explain variation in forecast accuracy. I produced a final model with a reduced set of variables that eliminates multi-collinearity to determine the magnitude and direction of the most important variables related to forecast accuracy.

Time variables

To determine how planning and forecasting practice and forecast accuracy have changed over time, I selected two different variables to describe the time at which a forecast was prepared. First, I used the year a forecast was made as a continuous variable to test for gradual changes that occur in the same direction over time. Next, I used a categorical time-period variable to indicate which of five federal laws authorized capital funding for transit projects at the time the forecast was made.[1]

---

[1] The five acts that authorized capital funding for rapid transit projects during the period of this study were the Urban Mass Transportation Act (1964 – 1987); the Surface Transportation and Uniform Relocation Assistance Act (STURAA) (1987 - 1991); the Intermodal Surface Transportation Efficiency Act (ISTEA) (1991 - 1997); the Transportation Equity Act for the 21st Century (TEA-21) (1997 - 2005); and the Safe Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users (SAFETEA-LU) (2005 – 2012). Since the sample is limited to projects that were completed before 2012, no forecasts completed under Moving Ahead for Progress in the 21st Century (MAP-21) (2012 – 2015) and the Fixing America's Surface Transportation Act (FAST-Act) (2015 to present) are included in the analysis.

Accuracy variables

All forecast accuracy variables are calculated as symmetrical percent error, as shown in Equation 2, where f is the forecast value and a is the actual value. Symmetrical percent error is recommended by Armstrong (1978) because it is dimensionless (it is scaled to allow comparisons of estimate accuracy for projects of different magnitudes) and symmetrical (it has the same range of possible values for underestimates as for overestimates).

Equation 2:

$$\frac{(f-a)}{\left(\frac{1}{2}\right)(f+a)}$$

For individual observations, symmetrical percent error will be positive for overestimates of actual values and negative for underestimates. When symmetrical percent errors are averaged across observations, the resulting symmetrical mean percent error (SMPE) represents the average *bias* of the set of estimates, but can overstate the average *accuracy* of the estimates, since any underestimates can mitigate the impact of overestimates (and vice versa) on the average value. Symmetrical mean absolute percent error (SMAPE) can better represent the average inaccuracy of a set of observations by taking the average of the absolute value of the percent error for each observation. However, SMAPE does not indicate the direction of the error. Since underestimates and overestimates have very different practical consequences in the case of cost estimation and ridership forecasting, this paper primarily focuses on SMPE rather than SMAPE.

Time variation in cost and ridership forecasts

Project sponsors typically update cost estimates at each project milestone (entry to Preliminary Engineering, entry to Final Design, and the execution of the Full-Funding Grant Agreement). Since information about the project itself and surrounding conditions improves over the course of project development, later cost estimates are often (though not always) more accurate than earlier cost estimates.

Flyvbjerg et al. (2003) and Pickrell (1992) have argued that the earliest cost estimates prepared during the project development process are particularly relevant to an evaluation of the potential influence of cost estimate accuracy on decision-making, since local decision makers choose to pursue or abandon a particular alternative based on those early estimates. However, the FTA's final decision to fund a project is based on the cost estimate current at FFGA execution. It has not been unusual for a project to advance from the Alternatives Analysis stage to the Preliminary Engineering stage, but never advance through Final Design to receive an FFGA (or at least not without substantial changes in project scope). Thus, the cost estimate prepared at each project development milestone has an important influence on whether the project will ultimately be constructed as planned. Moreover, the explanation for error in an initial cost estimate (which includes unforeseen costs that arise during project development and construction, including both cost escalation and cost overruns) must include factors that would not be included in the explanation for error in a final cost estimate (with includes only cost overruns, or unforeseen costs that arise during construction). Thus, in this dissertation, I evaluate the accuracy of both initial and final cost estimates. For five projects in the sample (Los Angeles Red Line, Salt Lake City University Line, Salt Lake City Medical Center Extension, Hudson Bergen MOS I, Hudson Bergen MOS II), initial cost estimates that correspond to approximately the same project scope as the project that was ultimately completed are not available. Thus the sample size for the accuracy of initial cost estimates is 62 projects. Likewise, there are ten projects in the sample (those for which data were taken from Pickrell (1989)) for which final cost estimates are not available. Additionally, for purposes of evaluating final cost estimate accuracy, the Los Angeles Red Line is divided into three separate segments. Thus the sample size for the accuracy of final cost estimates is 59 projects.

Occasionally, project sponsors will also update ridership estimates over the course of project development; however, it is common to carry the initial ridership forecast through to subsequent project development milestones. Thus, in this dissertation, I only evaluate the accuracy of initial ridership forecasts. For three projects in the sample, (Atlanta Heavy Rail, Seattle Transit Tunnel, and Denver I-25 Transitway), relevant ridership forecasts are not available. Thus, the sample size for ridership forecast accuracy is 64 projects.

Time variation in observed cost and ridership

One possible source of error in capital cost estimates is misestimating the effects of inflation, either because actual inflation rates are different than anticipated or because the time to project completion was shorter or longer than anticipated. For example, in 1970, the Washington Metropolitan Area Transit Authority estimated that a fourteen-month delay in the groundbreaking for the initial heavy rail system in Washington DC resulted in $55.3 million dollars in additional inflation costs (Schrag 2008), and Schrag (2008) suggests that the total inflation cost of delays on that project likely came to hundreds of millions of dollars.

The effects of inflation could be removed from a calculation of cost forecast accuracy by using inflation-adjusted values for both estimated and observed costs. However, this would require detailed information about the assumed rates of inflation and expected expenditure schedules used for preparing cost estimates, and this information is generally not readily available for cost estimates that have been prepared in terms of year-of-expenditure dollars. Moreover, accurately anticipating inflation and project schedules is an important component of preparing cost estimates and budgets. These factors suggest that the effects of inflation ought to be included in an evaluation of cost forecast accuracy. Thus, I have calculated cost estimate accuracy using nominal, year-of-expenditure costs for both forecast and observed costs. In some cases, year-of-expenditure forecasts were not available and midpoint construction-year values are used as an approximation.

Actual capital costs are evaluated at a single point in time upon project completion. However, actual project ridership (measured as an annual average of weekday boardings) varies over time, and measures of ridership forecast accuracy depend on the year selected in which to measure actual ridership.

Flyvbjerg (2005) reports that a common criticism of his work on the accuracy of demand forecasts for transportation infrastructure projects has been that opening year demand measurements do not capture the “ramp-up” in ridership that theoretically could occur in the first few years after project opening. In response, he argues that, while it may be preferable to use demand observations from a later year, opening-year data are typically more readily available. He further argues that opening-year forecasts should not include demand that is not anticipated to materialize until years after project opening.

In response to concerns with using opening-year data to evaluate the accuracy of ridership forecasts in required Before and After Studies, the FTA requires that comparisons between forecasts and observed data be based on observations taken after two years of revenue service (Federal Transit Administration 2000c). They explain,

> FTA recognizes that this evaluation will provide only a short-term "snapshot" of the performance of a new fixed-guideway system, and that many of the benefits, particularly in terms of land use, are long-term in nature. Project sponsors are of course encouraged to continue their data collection efforts beyond the period two years after opening. However, given the nature of the appropriations and authorization process, there is also a need for short-term data to provide an initial indication of the benefits of a project. (FTA, 2000, 76866)

While ridership-ramp-up has been discussed in general terms as described above, it has not been empirically studied. For this analysis, when data sources report ridership observations for multiple years, I calculate the accuracy of ridership forecasts using the actual ridership for the year following project completion that is closest to the forecast horizon year.

Variables describing planning and forecasting practice

As shown in Table 4, I use four types of variables to describe planning and forecasting practice: those that describe the elapsed time between the forecast and measurement of actual values; those that describe the financial characteristics of the project; those that describe physical project characteristics; and those that describe local experience with transit planning and construction.

The time between forecasts and measurements of actual values can be divided into three components: the time between the forecast and the beginning of construction (project development); time between the beginning of construction and project opening (project construction); and the time between the project opening and the year in which an actual value is observed (for cost forecasts, this value is always zero). To describe each of these three components, I use variables for project development duration, project construction duration, and forecast horizon (time from the year the forecast was made to the forecast horizon year).

For the Los Angeles Red Line, the project scope changed dramatically after construction began, requiring the preparation of revised forecasts. The project development time associated with such forecasts would be negative in this case. For purposes of this analysis, any negative project development durations were truncated to zero.

In many cases, observed ridership data were not available for the ridership forecast horizon year. In these cases, ridership observations from the closest available year were used and the difference between the year of the ridership observation and the forecast horizon year was included in the accuracy regression analysis as a control variable.

The analysis in this dissertation includes two variables to describe the financial characteristics of each project. First, it includes the actual total project cost. Second, it includes the share of total project costs that are funded through federal programs (primarily, but not exclusively, the New Starts program). The latter value is calculated as the final value of the grant awarded to the project divided by the actual project cost. It is worth noting that the value of a grant is generally determined to achieve an intended federal share, based on estimated project capital costs. However, the amount of money awarded does not automatically adjust if the project is completed over or under budget. When a project is completed over budget, this can reduce the actual federal funding share; when a project is completed under-budget, this can increase the actual federal funding share. Beginning with the passage of SAFETEA-LU in 1997, project sponsors have been permitted to keep a portion of the cost savings when a project is completed under budget (Duff et al. 2010).

Since some of the variation in federal funding share is directly related to the cost estimate error, the inclusion of federal funding share as an independent variable in the model regression models predicting cost estimate error would introduce endogeneity. Thus, I include federal funding shares in the regression analysis for ridership forecast error, but not for the cost estimate error models.

Two variables describe the physical characteristics of the project: project length and project mode. Project length is determined as the number miles the project covers in a single direction. Project mode classifies projects into one of six possible modes: Light rail, heavy rail, commuter rail, transitway, downtown people mover, and bus rapid transit. Transitways are defined as projects that create a dedicated right-of-way for transit without adding new transit service or transit vehicles. All transitways in the sample are bus lanes or bus tunnels. Bus rapid transit projects generally include the construction of bus lanes, but they are distinct from transitway projects in that they also add new rapid bus service.

Two variables indicate the prior experience of the project sponsor with respect to transit planning and forecasting. The first is an indicator variable for whether the project represents an initial line for a new transit system, an expansion of an existing system, or a renovation of an existing line. The second is share of the planned system mileage the project represents. For expansions of existing systems, this value is calculated as the project length divided by the sum of the project length and the total system mileage at the time the forecast was made. This value is always 100 percent for initial lines (since the total system mileage is zero before project opening) and zero for renovation projects (since these projects do not add mileage to the system).

Control variables for external conditions

External conditions, such as economic or demographic characteristics of the community a project is intended to serve, might influence the accuracy of cost and ridership forecasts, since these are important inputs to many forecast models. Moreover, project sponsors serving larger populations may have greater resources to devote to preparing rigorous forecasts. They may also answer to a wider variety of stakeholders, which could influence the incentives for promoting a particular project through optimistic forecasts. To account for some of these effects, the regression model predicting cost forecast accuracy controls for the population of the primary county served by the project sponsor and the unemployment rate in the central city of the projects sponsor’s service area in the year the forecast was made. The model predicting ridership forecast accuracy controls for four additional variables. Three of these represent the percent change between the year a forecast was made and the year the actual value is observed: change in the average price of gasoline within the state where the project is located, change in the central city unemployment rate, and change in the county population.

As discussed previously, ridership observations were not always available for horizon year referred to in the forecasts. In some cases, data are not available because the project was not completed until after the forecast horizon year. In others, it was because the system had expanded further by the time the horizon year was reached, and route or sub-route level ridership was not available in the forecast horizon year. To account for discrepancies between forecast horizon years and observation years, I include a control variable to indicate the number of years between the ridership forecast horizon year and the year in which ridership was observed.

Threats to validity

The quantitative analysis relies on a small sample size of just 67 projects. However, this is a relatively large sample compared to that of other studies that have empirically evaluated the accuracy of forecasts prepared for rail transit projects. Flyvbjerg et al.'s (2003, 2005) studies of cost and demand forecast accuracy of transportation infrastructure projects included 27 rail projects (including urban rail, high speed rail, and conventional rail), with the rest of the 210-project sample comprising road projects. Pickrell's (1989) initial landmark study on this topic included ten projects. Later FTA studies (Spielberg et al. 2003; Lewis-Workman et al. 2008) that updated Pickrell's (1989) analysis included all New Starts projects that had been completed to date. This study expands upon both of those samples through the inclusion of all thirteen additional projects that were completed after 2008 and for which Before and After Studies have been published.

Table 5. Results of *a priori* power analysis for regression models

	Sample size	Number of independent variables	Desired significance level	Desired statistical power	Required effect size	Required R^2 value
Type of forecast	**Full models**					
Initial cost estimate	62	19	0.05	0.80	0.45	0.31
Final cost estimate	59	19	0.05	0.80	0.49	0.33
Ridership	64	25	0.05	0.80	0.54	0.35
	Final models					
Initial cost estimate	62	13	0.05	0.80	0.36	0.26
Final cost estimate	59	13	0.05	0.80	0.38	0.28
Ridership	64	15	0.05	0.80	0.39	0.28

The primary concern associated with a small sample size is low statistical power, where power refers to the probability that no statistical relationship will be found when an actual relationship does exist (in contrast to a p-value, which is the probability that a particular relationship appears as a result of random chance, rather than being an indication of an actual relationship). Table 5 shows the R^2 values that would be required for each full and final model to achieve a significance level of 0.05 and statistical power of 0.80. The full models and final models presented in Chapter 7 all have R^2 values that exceed the thresholds indicated in Table 5.

Two sources of selection bias represent additional potential threats to validity of the quantitative analysis. First, forecast accuracy can only be calculated for projects that were selected for funding. Thus, forecasts for projects that did not receive funding are not included in the analysis. As discussed in Chapter 3, Eliasson and Fosgerau (2013) demonstrated that when forecasts are used as a basis for project selection, forecasts for the set of projects that is selected always will be optimistically biased, even if the full forecasts for all projects (including those that are not selected) would have been unbiased. The selection bias in favor of optimism may be mitigated if project evaluators reject some applications because of unrealistically optimistic forecasts, so that extremely optimistic forecasts are excluded from the sample as well.

Second, projects can only be included in the analysis if they have already been completed. Recent forecasts for projects with long project development and/or construction durations are therefore less likely to be included in the sample than recent forecasts for projects that were delivered more quickly. As a result, an analysis of trends in project development and construction durations over time overestimates the degree to which projects that have been started in recent years have been completed more quickly than projects that were started earlier in the history of the New Starts program. Thus, results referring to changes in practice that have occurred only in more recent years are less reliable than results suggesting consistent trends over longer periods of time.

Conclusion

Each individual project included in the study sample has a unique history with important contextual factors that are typically neglected in a quantitative analysis of the sample as a whole, based on data that are available for all projects. However, some of the richness that is lost in the quantitative analysis can be added through qualitative analysis of interviews with transit professionals who have worked within the framework of the New Starts process throughout the history of the program. Taken together, the interviews and the results of the quantitative analysis can offer important insight into how changes in the New Starts program have led to changes in the perspectives and practices of project sponsors over time, and how these changes have in turn led to changes in the accuracy of cost and ridership forecasts for fixed-guideway transit projects.

Chapter 5. Perceived Purposes and Consequences of Federal Policy Changes

The purpose of the interviews was to gather perspectives from transit professionals on how changes in the New Starts program had influenced transit planning and forecasting practice over time. In describing their experiences with the New Starts program, the practitioners whom I interviewed for this study described changes they had observed in fixed-guideway transit planning over the years, as well as changes in the federal policies governing the New Starts program. They also offered their perspectives on the reasons behind changes in the New Starts program and the intended and unintended consequences of those changes.

The consequences interviewees described can be grouped into three categories: First, there have been adjustments to redefine the federal role in funding capital transit projects. These types of consequences were more commonly mentioned when interviewees discussed changes that occurred in the earliest years of the New Starts program. Second were changes that were perceived to —intentionally or unintentionally— affect the fairness of the competition among local project sponsors for federal funds. And third, in discussing the most recent changes to the New Starts program, interviewees were most likely to mention the way in which the program has improved local planning practice. I summarize and discuss comments in each of these three categories in the sections below.

Defining the Federal Role in Transit Funding

Some interviewees recalled that, in the earliest years of the New Starts program (particularly during the Reagan Administration), when the federal government declined to fund local transit projects their refusal was perceived as being connected to a broader ideological commitment to devolution. As the New Starts program matured, the explanations for changes to the program offered by interviewees shifted from a focus on small government to an overall failure of the federal government to generate sufficient revenue for infrastructure in general. This failure to generate funding for transit capital projects was not necessarily interpreted as a lack of interest in the outcomes of such projects, since the perception of and increased scarcity of federal funds was accompanied by an increased federal role in providing technical assistance and oversight.

One interviewee who had worked for the UMTA during the Reagan Administration described how, in the early years of the program, the perceived purpose of New Starts rating criteria was to limit the federal role in transit funding. Prior to 1984, without explicit criteria to evaluate projects, federal funding for capital transit projects was allocated primarily through Congressional earmarks. The New Starts project evaluation process represented an effort to shift the power from the legislative branch to the executive branch to allocate, and thus limit, New Starts funds.

> Reagan's first budget said, "We're going to do away with the New Starts program entirely. We don't need that. It's not a federal responsibility." And so for a while, we [UMTA staff] sort of kept our heads down and hoped that they didn't get chopped off. But by 1984, the administration had come to recognize that Congress wasn't going to let the program go away. Congress was continuing to earmark the program and saying, "Yes, President Reagan, you shall fund this program and here are the projects you shall fund." And they would earmark money and the administration was then obligated to make those grants. So after three years of that, the administration said, "You know, if we're going to have a program, we want to manage the program. Let's create a ratings system. We are going to fund these projects, but only those that are the best in terms of cost effectiveness and in terms of local financial commitment." Through the remaining years of the Reagan administration, they never really had their heart in funding them at all. — *Interview 4*

A consultant likewise recalled the perception in those early years that some of the federal policies in place at that time were intended to find reasons not to fund projects. In response, Congress began to proactively specify the criteria by which the UMTA, and later the FTA, could evaluate proposed projects.

> I think [UMTA staff] tried to do the right thing. Their hearts were in the right place. They were viewed as the guys that always said no, but they were just following the rules. A lot of this was during the Reagan administration where the whole intent was to kill the projects through the New Starts process and not have any in the budget. Some of the ups and downs of the program have come as Congress has changed the rules and the laws. I can't imagine too many

> other programs where Congress has really dug down into the nitty-gritty to define those requirements. — *Interview 11*

The federal role in local transit projects extends beyond simply funding the projects to assuming a portion of project risk and providing project oversight. From the early years of the New Starts program, observers, notably Pickrell (1989, 1992), argued that local project sponsors should bear a greater share of the risk of cost overruns. Indeed, an interviewee who had worked at multiple local transit agencies described the current trend in the New Starts program toward transferring not only funding obligations, but also risk from the federal to the local level. The interviewee expressed his view that this transfer of risk was not necessarily intentional, but was rather a by-product of an effort to streamline the project development process.

> FTA has done a lot to streamline the process, and I think what they've done is transfer the risk. It's the same amount of risk and the same amount of work, but instead of sharing in it equally, it's been transferred [from FTA] to the project sponsor earlier in the process. So on the surface, it looks like the feds have streamlined things, and they have in many areas. But how it's playing out on the streets is that the same amount of time and resources are necessary to advance a project. You really have to have your financial ducks in order to move a project forward, which would be very difficult for agencies that aren't already set up that way from doing previous expansion projects … because a lot of the risk and the onus would be on the agency, whereas before it was shared. — *Interview 10*

In contrast, an interviewee who had spent time both as a part of the FTA staff and as a consultant to project sponsors expressed the view that shifting risk from the federal to the local level has been both intentional and appropriate, although it has not been popular with project sponsors.

> FTA implemented a policy last year that was very unpopular with the industry: When a New Starts project advances from Project Development to Engineering, the sponsor must be willing to lock in the amount of New Starts funds that they will request. And the industry said, "Oh, it's too early in the process for us to know how much we really need!" But to me, it's a great example of FTA transferring risk to the grantee. My feeling is that if the project sponsor does not have the confidence in knowing how much they should ask for, they're not ready. They're not ready to take that step. That puts the pressure on the project sponsor. Of course, the industry is all project sponsors, and they really disagree with that particular policy, but I think it's great. I think it puts the

> onus on the sponsors to carry risk. If they're not willing to do it, then they're not ready to advance. — *Interview 6*

A transit agency employee also referred to the increased burden on local project sponsors resulting not from reduced federal involvement in funding, but from increased federal involvement in project oversight. As the reporting and oversight burden for New Starts projects has increased, the interviewee's agency has begun to fund its smaller, more straightforward projects locally and apply for New Starts funding only when the project is so large and complex that the federal contribution would represent enough money to warrant the additional effort.

> So when we go through the New Starts program, they assign a project management oversight consultant … to look over our shoulder and monitor what we do. And that project management oversight consultant usually will … do monthly reports, they have monthly meetings with us, they do reports to Congress and the FTA on how we're doing, and they evaluate what they think is going well and what they think are areas of high risk. And so we have to spend a lot of time and effort working to satisfy all of those reporting requirements when we're in a New Starts project. When we spend our own money and don't have federal dollars … we don't have oversight consultants like we have under the New Starts program. That's why we reserve New Starts for our biggest, most expensive projects, … We can actually get more money when we put our biggest projects forward. … There's a lot of extra work involved in going through the Federal funding programs. And so we want our efforts to be put into the projects that have the biggest return for us. — *Interview 12*

A consultant described how FTA oversight has imposed a degree of standardization on the development of models that might otherwise be more customized to reflect local observations and context.

> FTA is very sensitive about what the values used in those mode choice model utility equations are. … We want to be sure that [the coefficients in] our mode-choice model will meet their targets. So rather than take our chances and estimate a model and say, "Oh no! The estimated data doesn't fall into the range specified by FTA," we assert the coefficients and then we'll update as necessary to calibrate. But we begin with a model that FTA has already approved. — *Interview 13*

In the view of some interviewees, much of this additional oversight is a direct consequence of the research by Pickrell (1989), which is perceived to have damaged the reputation of the New Starts program and motivated a sharp increase in project oversight in response (although the establishment of the PMO program preceded the publication of the Pickrell report by a couple of years).

> When the program first started having projects open … the Pickrell report documented some really horrendous misses, which is a reputation that the program has struggled to shake in all the years that have gone by since. — *Interview 8*

> I remember the Pickrell Report back in the 1980s. … [In response], the Feds have said, "Alright, if the concern is that travel forecasting isn't being done right, we're going to have to spend a lot more time looking at it. If it's not, then it's project management. We're going to put in a project management oversight consultant. If you aren't dealing with risk appropriately, we're going to require you to do risk analyses and Monte Carlo processes to make sure you're getting the right cost for a project, that you're staffing it right, and that you're just not as optimistic." So I think FTA has gotten more rules in as an equal and opposite reaction to other people coming in and being wide-eyed and rosy-colored glasses. — *Interview 1*

> There was a period in the 80s and the 90s when a lot of projects were way over budget, and that really undermined congressional confidence in the program. And I think FTA has done a great job in improving the oversight of projects so that there are fewer projects being delivered over-budget or with massive delays. It still happens —these are big infrastructure projects— but it doesn't happen as much. — *Interview 6*

More recently, the perception that federal regulations limited the federal role in transit funding out of an ideological commitment to small government has been supplanted by an emphasis on the overall scarcity of federal money for any purpose. An interviewee described the failure of the federal government to play a larger role in funding capital transit projects as part of a larger failure to fund infrastructure in general.

> Getting any infrastructure projects funded by the federal government is a major task these days. We have decaying infrastructure all over the country, not just in transit, but also in highways, bridges, tunnels, water, and sewer. Any kind of infrastructure is woefully underfunded in the country today. We're very grateful for any funds we get through the New Starts program. We hope that Congress will become a little bit more aware of not letting our infrastructure decay in the coming decades and will start putting more money into it, but in the meantime, we feel that [our local jurisdiction] is kind of on its own. … Thirty years ago, the federal funding share was eighty percent and the local was twenty percent. Then it shifted more to fifty-fifty, and now we're at about two-thirds local and one third federal. So we've seen this shift progressively from the state and federal government to the local. We're grateful for New Starts; we would like to see the federal government step up and fund more of these types of programs, and in the meantime, we're trying

> to make our case [locally] that voters need to step up and fund a larger share than historically they've had to do. — *Interview 12*

One interviewee described ridership forecasts as a mechanism by which FTA can limit the federal role in transit, regardless of whether the need to limit the federal role stems from ideology or financial constraints.

> The big thing [the FTA has] typically been holding us up on is ridership. That's been their governor, if you will, like the governor on an engine that limits how much you get through. That's been their tool to limit how many projects are getting federal funds. — *Interview 2*

When project sponsors perceive that the FTA is generally disinclined to fund any transit projects and is looking for reasons to deny applications (as a few interviewees suggested has occasionally been the case), they may come to believe that the FTA cannot be trusted to fund necessary and cost-effective projects based on accurate forecasts of costs and ridership. Tetlock (2009) has observed that, when experts do not trust decision-makers to place adequate weight on their forecasts for the costs, benefits, and risks associated with a proposed action, they may justify presenting forecasts that represent best-case or worst-case scenarios in order to lead decision-makers to an optimal decision. If cost estimators and ridership forecasters adopt this strategy, we would expect to see more optimistic forecasts at times when the FTA is perceived to be antagonistic toward funding transit projects in general.

Fairness in Project Evaluation and Comparison

With the recognition that available federal funding is very limited has come an increasing awareness on the part of local project sponsors that they must compete with one another for these scarce funds. This awareness is accompanied by a concern that the competition should be fair.

> Number one, we all recognize that FTA has a limited budget, so the reality is, part of the New Start process is to get the best projects. … There's a very challenging exercise to compare one transit property to another. So the essence of it is a fair effort. — *Interview 5*

Interviewees explicitly referred to the application for New Starts funding as a competition with other cities and agencies that are competing for the same funds.

> The competition is getting fiercer right now because many more cities are competing for federal New Starts funds, and the pot of money that they're giving out is not increasing dramatically. Congress has had fiscal difficulties for any kind of infrastructure project in the country. Congress tends to be deadlocked over funding new infrastructure and hasn't really been able to approve a lot of new funding measures to support transit and highways and other types of transportation. But the cities' demand for these projects is going up and a lot of cities … rely upon federal funds in order to build a project. As a result, there's just a lot of competition now for these federal funds and a lot of cities are just going to be disappointed. They're not going to be awarded grants. When we choose to go after federal funds, it's a major effort because we have to make the case for why our project … should be funded over other cities' projects. — *Interview 12*

When the application process is framed as a competition and project sponsors must compete on the basis of their demand forecasts and cost estimates, they are especially concerned that all competitors are evaluated fairly. In talking about the project evaluation process, many interviewees used analogies to competitive sports and games, with references to playing by the rules or leveling the playing field.

> The job of the folks in Washington is to make sure that everybody is playing by the same rules. When all the different agencies go to Washington and one claims, "Well, we're going to get fifty thousand new boardings by our project," and somebody in another city comes in and says, "Well, we're going to get seventy thousand," I think everybody wants to make sure that everybody is using the same rules in doing the ridership modeling. Everybody needs to play by the same rules so that somebody can't claim a higher forecast and get awarded funding on the basis of that if they didn't use the same rules. So I think FTA's in this position where they've tried to level the playing field and make sure everybody is using the same methodologies. — *Interview 12*

> The federal agency that has this program that only gets a limited pot of money each year, and they have requests for this money coming in from all over the place for all different types of projects of varying quality and usefulness in terms of how much the project will actually solve a problem that needs to be solved. … They have to have a set of standardized processes, so when you get a set of cost estimates — one from City X, another from City Y, and another from City Z — they've all been developed in a common methodology so you can look at the cost estimates or the ridership forecasts and can evaluate them based on numbers that are not distorted by methodological differences, or differences in the level of rigor. So you try to explain that to your clients that this is not a 'gotcha' type of thing. It's in their best interest to understand where [the FTA is] coming from and go by that game. And I don't say 'by that game' in a

> negative sense, but they can help their case by playing within [the FTA's] framework. — *Interview 11*

The analogies to a game with consistent rules and a level playing field refer to fairness of process. In addition to fairness of process, some interviewees also referred to the perceived fairness of geographic outcomes. For example, a few described how the FTA avoids concentrating New Starts projects in a small number of cities, even if those cities hosted the lion's share of high-performing projects. An interviewee who had worked at FTA described efforts to ensure that projects were geographically distributed as politically driven, but that those considerations were secondary to an objective evaluation of project merits.

> The evaluation process is rigorous and very much merit-based. Projects that were highly recommended … really had merit. Where politics comes in is when you've got a slew of projects, all of which rate well, but there's not enough money to fund them all. Then it becomes a political question and it's often been driven by trying to distribute funding across the US. … If a particular region or state has multiple projects that evaluate well and are ready for a grant, sometimes only one [of those projects] might be selected instead of the two or three that would qualify. But there are exceptions to that. — *Interview 6*

Political concerns with the spatial distribution of funded projects arise at the local level as well as the national level. One transit agency staff member described how the decision to propose a project for New Starts funding incorporates considerations of perceived geopolitical equity at the local level in addition to expected project performance.

> We have [many] members on our board that represent all areas of [the region] and usually a project in New Starts is only in one of their areas. … If we say we're going to build a New Starts project in this particular part of [the region], we'll have [the rest of the] board members say, "Well, why aren't you building it in my area?" We have to come up with a long-range transportation plan that has something for everybody in it. It has to be geographically distributed and provide something for everybody. Then we have convince all of our board members that they should go after New Starts funding in one area because if we get funds for New Starts here, it might free up more local funds in another area for a project that might be in their district. So there's always this balancing of not just project performance, but how does that balance out with other projects that might be serving other areas so that everybody feels it's a fair thing to allow this project to go forward into New Starts and not something in their district. So there's always a political part of it too that the board has to consider that not one part of [the region] is getting all the new transit benefits. — *Interview 12*

A consultant likewise described the role of project evaluation criteria in ensuring that funding is distributed among various geographic regions, across modes, across projects of different scales, and for projects that are perceived to provide different types of benefits.

> [The FTA has] taken measures to make the evaluation of transit projects and investments more equitable between large and small projects and multiple types of rail initiatives. … They've also diversified the portfolio of investments geographically. There was a period in the early to mid-2000s where all the New Starts money was really in the south and west, and that was the unintended consequence of the calculation around transit system user benefits, which measured project benefits based upon how much time you can save, and not necessarily on the project's true cost efficiencies. The heightened role of other factors besides cost and ridership has contributed to that diversification as well. They strengthened the role of land use and economic development, which is huge. The steps that they took when they implemented the environmental benefits, mobility benefits, and congestion relief have certainly contributed to FTA being able to implement a more diverse array of projects. The types of benefits that these new transit projects have are just much more robust than they were before, because of this diversification of investment. — *Interview 10*

Cost effectiveness has been a primary measure used in project evaluation, but the methods of calculating cost effectiveness and the weight assigned to this metric have changed over the years, as described in Chapter 2. One interviewee described these changes as a response to concerns on the part of project sponsors about the methods' fairness.

> [Cost effectiveness metrics] have changed as, really, the industry has given comments to FTA, saying, "Well, this is not quite fair" or "That's not quite fair." — *Interview 1*

Of course, fairness is often largely in the eye of the beholder. One agency staff member expressed the perception that any change in the method for measuring cost effectiveness will benefit some cities at the expense of others.

> The measure for cost effectiveness has changed over the years. … Any time they change the rules, different people get hurt with it. Right now, it's the number of riders on the high-capacity mode, which really helps a city like a San Francisco or Chicago or New York. They have just a ton more riders. It may not help an early city that's trying to develop a light-rail system because they may not have the same ridership. It used to be the travel-time savings for all the people that were associated with the mode. … To some degree, it changes how you think about a project. — *Interview 1*

Another agency staff member further described how changes in cost effectiveness measurement over the years have influenced the types of projects that are eligible for funding, and expressed a belief that the shift from a simple measurement of cost per new rider to TSUB benefitted projects that served longer trips, such as commuter rail, and that a later emphasis on current ridership has benefitted projects in systems that already have an established base of transit riders.

> When I started working in project development, it was a pretty simple calculation of cost versus ridership. At the time, the FTA focused on cost per new rider, and you'd be comparing horizon-year ridership, so it was cost per new rider for 2040. … Then FTA changed that to look at user benefits. And user benefit measured whether there were travel-time savings that happen from the project. When they went from just riders to that travel-time saving measure, it really changed the kinds of projects that could qualify for New Starts funds. It really benefitted long-haul light-rail projects. It benefitted commuter rail projects. Streetcar projects didn't really show particularly well because they're not really saving anybody travel time. … [At that time], benefits were still based on a horizon year. … With this latest shift, they're more focused on existing corridor ridership. That's really hurt our ability to compete for federal funds. Because most of these corridors we're looking at for development, there's not a lot of existing service in those corridors. … So this new focus on immediate ridership has really disbenefitted a lot of our projects. Past projects … that were federally funded … would no longer qualify for federal funds. With this latest change if you look at the projects that are now competing for federal New Starts dollars, it's largely BRT projects because their cost-effectiveness thresholds are so high, and their existing ridership thresholds are so high. It's screened out a lot of light-rail projects and really focused in on BRT. I don't know if that was the intended consequence of the federal rule change, but that's been the outcome. — *Interview 2*

An interviewee who had closely observed the discussion around changes in project evaluation criteria confirmed that the shift away from including travel time savings in calculating project benefits was not an attempt to influence the type of projects or types of cities that should receive funding, but rather to move back to a simpler, more understandable metric.

> I think the switch to user benefits was significant because it did try to capture all of the transportation benefits of a project, not just new riders. But in part it was the seeds of its own undoing because it got really complicated and it required sophisticated modeling. … In MAP-21, Congress said "Enough of that! That's too complicated!" I think this administration was also trying to step

> away from that kind of a measure. We ended up with cost per project trip, which to me is a step backwards because I don't think that measure is a particularly good indicator of benefit at all. You can have a lot of people riding on a project, but are they better off? — *Interview 4*

The transparency and understandability of a cost-effectiveness metric also relates to fairness, since understanding and producing a complex metric may require expertise and resources that are not equally available to all project sponsors or to members of the public who wish to participate in decision making. One transit agency staff member described how TSUB was too complicated for the general public, and even for agency staff to understand.

A consultant explained that, although there may have been a perception that changes in cost-effectiveness metrics influenced the types of projects that would be eligible for funding, that the projects that performed well by one measure generally performed well by the one that replaced it.

> We did a little exercise to see how cost per hour of user benefit (the old measure) correlated with cost per project trip, just to see if it really changed the playing field. To my surprise, it really didn't. The ones that were good under the old measure are still good under the new measure. So maybe it's okay. — *Interview 4*

However, even the perception that a particular cost-effectiveness metric may favor a certain type of project can influence whether projects of that type are proposed in the first place.

> If you look at what could qualify, and what came out of those three different periods, they're quite different projects. And those FTA criteria end up getting echoed forward in the long-range plan. The MPOs say, well, unless it can qualify for federal funds, we're not going to include it in the long-range plan. And so by *de facto*, they're really defining the projects that you end up pursuing, and I don't know if that's their intent. — *Interview 2*

Like the inclusion of travel-time savings in calculations of user benefits, the comparison of projects to a TSM alternative was viewed as offering a marginal improvement in evaluating a project's true benefits, but at the expense of creating a process that the general public could easily understand.

> Before, we were doing this cost-effectiveness equation, but it was in reference to this theoretical transportation system management alternative, which was basically a really optimized bus project. It was a surrogate for "Okay, what are the benefits of this project above and beyond what you could do with a much less capital-intensive [project]?" … Having that not be a requirement anymore is a huge, huge change. It was a really confusing time in describing the New Starts process to the public. Because they didn't understand – what is this TSM project? We have an alternatives analysis that we've done at the local level, but FTA would still require us to do this additional one with a different baseline referring to this TSM alternative. It just made things really complicated and challenging. But I think from [FTA]'s perspective, it was trying to force local jurisdictions to really evaluate whether or not they really needed to spend all this money to get these positive impacts. I appreciate the idea; I just think it was a clumsy way to do it. — *Interview 3*

Not all interviewees agreed that a the project evaluation criteria ought to be neutral with respect to which types of projects or types of cities that could receive funding. One argued that the current criteria may seem to favor bus rapid transit because that really is the right mode for most cities that do not already have a fixed-guideway transit system.

> I'd love to see more higher-end bus rapid transit. … It's been said, and I don't know if I agree with it, but somebody once said that all the cities that need urban rail systems have them. All the cities that need them have them. And now, these next-tier cities in terms of size should be thinking about bus rapid transit and lower-cost investments. And maybe you start with BRT and you build for the market and then you go to a higher-capacity mode of transportation. — *Interview 6*

The same interviewee described how the creation of the Small Starts program in 2005 created better opportunities for smaller cities to propose projects that were more suited to their needs.

> I think the most significant change in the law came about beginning with SAFETEA-LU when the Small Starts program was created. … I think it was great because it got people thinking about, "Hey, we want to do this million-dollar project, but we could never afford it, so let's recalibrate our expectations, and let's not swing for the fences and get nothing. Let's do something more modest." That something more modest [often] turned out to be bus rapid transit. … I think that helped BRT gain cache because all of a sudden, here is some credibility from the New Starts program to the Small Starts sub-program to advance the more modest projects more quickly, and therefore distribute the benefits of transit to more and more regions. So in regions that can't support a light rail system, but they can support bus rapid transit: here's a program that allows them to do that. … That's been the most significant change, and a very positive one. – Interview 6

Another interviewee described how the creation of the Core Capacity program offers funding for large cities with existing, congested fixed-guideway transit systems so that these more established cities do not compete directly with mid-sized cities that are not seeking to relieve transit congestion, but rather to expand and introduce service.

> You have huge metro areas like New York and Chicago and Boston and DC that have much more capital-intensive projects and huge benefits as well, so they've tried to add this Core Capacity change that I think is designated for those areas that have such strong ridership that they're always going to out-compete the Dallases, the Alberquerques, the Phoenixes, the Salt Lake Cities, and the Portlands of the United States. It's nice that they've now created different types of programs, so that you have a better chance of getting funded. — *Interview 3*

The creation of additional grant programs thus resulted in a pool of New Starts applicants with more comparable projects. As one interviewee describes, this relieved some of the political pressure to advocate for projects that might not be a good fit for the local context.

> There was a lot of naiveté on the part of transit properties. They were under political pressure, where if a project got funded in another city, local leaders would say 'we want one too.' I think today, there's been a separation of the big players from the not-so-big players, in terms of [competing for federal funds]. I think what's helped is that there are now these other programs like Small Starts and TIGER grants and so on. Those provide a discretionary funding mechanism for the mid-sized and smaller municipalities or metropolitan areas to do projects of an appropriate scope and scale for their situation. What you've got left are people who are going after New Starts funding for major capital projects. — *Interview 11*

Consultants may also face pressure from clients to produce optimistically biased forecasts, and they may seek to avoid a reputation for working on projects that are not approved for funding.

> They were interested in someone that would give them a forecast that would be high enough to justify the project that they had already decided that they wanted. If that's the criterion that people are using [to select a consultant], then they would be seeking somebody who could jigger up a forecast that would be high and they wouldn't be so worried about whether or not they achieved it. — *Interview 4*

> You never want to be perceived as the guy who killed a project. Because if you get a reputation that anything you touch kills a project, then you'll never get

> work. And that gives you all the insight I could ever want to give you about how the industry sees itself, and how cutthroat the industry is, and how much it's hard to really care about doing a reliable forecast. — *Interview 9*

However, the existence of other funding programs with different purposes allows consultants to shift their role from making sure a project gets approved under any particular grant program to helping a client choose the right grant program for a particular project. One consultant described a consultant's role in those terms.

> I perform a preliminary rating so they understand if they have a competitive project or not, and I help them evaluate if the New Starts program is a program that's right for that project. — *Interview 6*

Overall, fairness was described as being achieved through creating separate competitions for particular types of projects and by ensuring that all participants within a given competition played by the same rules. Preparing forecasts as part of a New Starts application was thus seen as one of the rules of the game, or as one consultant described it, "a hoop to jump through."

> They often, particularly in the early days, a client would want a light rail line, or they would want this or that project, and they viewed the New Starts process as a hoop to jump through in order to get the federal funds. — *Interview 11*

The perception of project evaluation as a competition supports a belief that forecasts should be prepared in way that they are honestly comparable among competing projects, but it does not necessarily require that forecast accuracy should be maximized, as long as inaccuracies are consistent across projects and thus do not affect projects' overall ranking. In fact, if project sponsors believe that overstated benefits and understated costs are the norm, then this could motivate them to produce optimistic forecasts in pursuit of fairness.

If the New Starts application process is viewed as a game, game theory may offer insight into the behavior of project sponsors. Hofstadter (1985) has proposed a version of the Prisoner's Dilemma originally proposed by Tucker (1983) called the "closed-bag exchange." Two individuals agree upon an exchange of items of equal value (money for diamonds, for example) that will take place by each individual placing the agreed-upon item in a closed bag, then exchanging the bags and parting ways before examining each bag's contents. Each individual can choose either to cooperate by following the agreement and placing the agreed-upon item in the bag or to defect by exchanging an empty bag.

If both individuals cooperate, they each end up with the same value of items they started with. This is likewise true if both individuals defect, since they both keep the same items they started with. However, if one individual defects and the other cooperates, the defector is better off after the exchange than before, and the cooperator is worse off. Thus, if one participant doesn't know whether the other will defect, and if there are no reputational effects of defecting, then the optimal decision is always to defect, since there is no scenario in which cooperating offers a better outcome than defecting.

If intentionally producing optimistic forecasts is a binary decision (to produce forecasts that are either as accurate as possible or as optimistic as possible while still plausibly credible), and the amount of credible optimism that can be introduced to a forecast is approximately consistent across project sponsors, then this decision can be thought of as a version of the Hofstadter's (1985) closed-bag exchange. If all project sponsors choose to produce accurate forecasts, the chance of a particular project being selected is no different than if all project sponsors choose to produce optimistic forecasts. However, if some project sponsors produce accurate forecasts and others do not, the projects sponsors who produce accurate forecasts are at a disadvantage relative to those who produce optimistic forecasts. Thus, game theory would suggest that the rational choice for a project sponsor would always be to produce optimistic forecasts.

Published information that indicates that cost estimates and ridership forecasts are usually optimistically biased could influence the behavior of project sponsors in two ways. First, it introduces a reputational cost when a project sponsor "defects" by producing a very optimistic forecast, and this may increase the incentive to produce accurate forecasts. It also signals to other projects sponsors that at least some project sponsors have gotten away with optimistic forecasts, and thus increases the incentive to likewise produce optimistic forecasts in order to "level the playing field." Such signals may relate to the broken window theory popularized by Wilson and Kelling (1982), which suggests that criminals react to signals (such as broken windows) that indicate the level of crime that is tolerated. Similarly, published reports indicating that projects that are widely regarded as successful were planned using optimistic forecasts may signal to project sponsors that forecast optimism is an acceptable strategy to get a project approved.

Improving Planning Practice

Some interviewees expressed dismay at the apparent widespread bias in forecasts produced for New Starts applications and described a strong sense of professional duty to produce accurate forecasts, regardless of the incentives to do otherwise.

> Victory is defined as whether we got the project built or not, but if you're just trying to get the project done, the ends don't always justify the means. If the forecaster has to lie to get the project built, it's not something I want to do. I've always thought that the forecaster has to have a higher sense of public duty or a higher calling. … To be frank, the way I see it is, I want to be a consultant, not a prostitute. And if we don't have a little bit of a higher calling, then we're going to be there to do whatever the client wants. That is part of our duty, but part of our duty is to the public as well. It's our duty to provide a fair forecast. … The strength of a forecaster right now is to give an honest forecast and to let the project stand on its own merits with a reliable forecast rather than trying to optimistically tweak it to make it go further than it should. — *Interview 9*

A few consultants shared stories about standing their ground in the face of client pressure to move a project through the project development process in spite of its poor cost-effectiveness.

> [A transit agency] brought us in saying, "We need your help to do a forecast to get this project funded." And I was quite proud of how we approached it. We said, "We'll give you a forecast that will be acceptable to FTA, but we can't guarantee that the project will get funded." We took the high ground and said, "We'll give you the best, most reliable, most credible forecast there is, and

> then the chips will fall where they may." I thought that was successful even though the project didn't get funded. … I always get some satisfaction from riding a project after it gets built, but we're doing a good bit of Before and After Study work now and finding that a lot of the projects that did get funded, that are now operating, they overran their cost estimate, or their ridership isn't what they expected it would be. I guess that's a legitimate question: Is that success then? If the forecast was X and ridership is only half X, is that a successful endeavor? — *Interview 4*

> There was a city … that called me in at the last minute as they were trying to get a full-funding grant agreement. They had already ordered rail cars, and I don't know how it went this far. They thought I was going to come in and make it all better. We looked at their travel patterns that they were using to make the case for the project, and the growing travel pattern was in the north-south direction, but they were proposing an east-west rail line. It wasn't anything to solve the problem they thought they had. The project went down the tubes. I sort of felt like the guy in the emergency room when they bring you a patient that's not breathing, their heart's not beating, and they've been decapitated, and they look at me and say, "You killed my child." I got the blame for killing their project. — *Interview 11*

> We put everything into the model and told them what we came up with. We said they weren't going to get more than 400 or 500 people on the rail line, and they were expecting thousands. The project was found not to be … competitive for the federal funds they were hoping to seek. That took a lot of fortitude to go into a project steering committee and tell them they don't have a project, … but we stuck to our guns. The thing I'm very proud of is that we actually turned the conversation around and told them, "If you really want rail, here are some things that you need to do to build up the market for rail." That goes into land use plans, building up your current bus system, … and making transit a real option. I was really proud of that because that was where the conversation needed to go. The conversation didn't need to go to "How do I get a two-billion-dollar rail line?" It should have been, "Instead of a two-billion-dollar rail line, why don't you make a policy choice to support transit?" — *Interview 9*

However, a transit agency staff member succinctly explained that investing any planning resources into a project that never gets completed, and thus delivers no benefits, is even less cost effective than completing a project that delivers only minimal benefits.

> If there is no project then it's not cost effective. Then there really is no project. — *Interview 3*

When a project sponsor believes that their project may only be marginally eligible for a New Starts grant, they may seek to make the project more attractive by modifying project characteristics. This might involve finding ways to lower a project's costs.

> When projects are kind of borderline, we look for ways to improve a project's rating. That could mean modifying the scope such that costs are reduced but benefits are not, or at least that benefits are commensurate with the costs. — *Interview 6*

> We're always keeping an eye on the prize, which is, "Is this going to be an eligible and a competitive project?" So as we're doing project planning, we're always trying to make sure that this project is generating the ridership and the usage that justifies the cost. And if it doesn't justify the cost, then we need to bring the cost down and make a project that is more in line with the cost-effectiveness equation. — *Interview 3*

Additionally, a project sponsor could modify a project to attract more riders. One interviewee described how simple changes like finding ways to improve vehicle speeds or increase park-and-ride capacity can be enough to improve a project's rating.

> If it's not a project that's going to rate well, then we need to do some tweaks. We might increase park and ride capacity or make it faster — whatever it takes to make a project that's competitive. — *Interview 3*

As an alternative to making changes to the actual project, a project sponsor may also modify the forecasting model specifications in a way that will produce a more favorable forecast. One interviewee mentioned this as a possibility, but emphasized that it should be done with effort to correct errors, rather than simply to inflate the ridership forecast.

> Despite our best efforts, sometimes there are errors. … As we're doing these projects, even though they take years to go through the planning process, it seems like every time we've got a submittal, or we need a decision made and we're putting together the data, things get rushed. And so, and it seems like every time we do a new model run, we find something that we were missing before. … They're not just tweaks, but they're catching omissions or errors. … Those are the types of quick things that should be done regardless of where you are, but just because of time constraints, you may not focus on them unless you're running into issues with cost effectiveness. You can only do so much on the model. It has to be as reasonable and reflective of reality as possible. You can't just tweak something to increase ridership purposefully. You have to find out if there's a problem. — *Interview 3*

Another interviewee explained that, while it's more common to improve a project's rating by changing actual project characteristics, it is possible to improve the rating by changing the model specifications, as long as model inputs are within a range that FTA determines to be reasonable.

> After we went through an exercise to try and increase ridership and reduce costs, we ended up changing the alignment, and then tinkering with the end of the lines as well as doing a complete cost refinement as well. And there were several things to it. That's usually where folks focus their efforts. You could also do sensitivity testing around some of the forecasting, either in the STOPs model or the your regional model. You can do different sensitivity tests with those. Sometimes they'll give you different answers. You can do tests with the projected year. There's lots of stuff you can do with the modeling effort. FTA has to buy off on it, but there are things you can do there. — *Interview 10*

Some of tension between presenting a project in the best light and producing an accurate forecast can be relieved when forecasts are used not only to make the case for a project's eligibility for federal funds, but also to help local leaders choose the right project in the first place. A consultant described explaining to project sponsors that the forecasts they produce are not only to help the FTA make comparisons among competing projects, but also to help the project sponsors make prudent decisions.

> They often want a light rail line because of this idea that you're not a major city until you have a rail transit system and a baseball stadium or a football stadium, so they would often have a solution before they defined the problem. It was often a solution in search of a problem. I try to explain that the FTA wants to make sure not only that they have a legitimate process to make recommendations for allocating a resource, but also that local decision makers should have a full understanding of what the problem is that they're trying to solve. … It takes the local decision makers to ask the same questions that FTA is asking. Don't just go through this process to make yourself eligible for federal funds. If you go through all this effort of planning, use it for your own information. We often stress that this process was to … provide the decision-makers with information to help them make a decision. — *Interview 11*

The same consultant described how the perspective of using the New Starts process to facilitate more informed local decision-making has gradually taken root over the years, and offered two explanations for why this has been the case. First, the elimination of earmarking projects for New Starts funding signaled to project sponsors that funding would be based on performance rather than on politics.

> In the early days, for a project to get New Starts funding, it had to be earmarked. Even though FTA would make a recommendation that would go into the president's budget, there actually had to be an earmark made for it. … There was always a feeling, because earmarking was a well-established practice and … that was how you got federal money for projects: It was through lobbying your congressional delegate. So the idea was that, if this is a

> discretionary program, it means you should go in and do the stuff that gets you an army base or a project under some other federal program. That's how it's done. It took a while to get people to realize this is not that type of federal discretionary program, and that this was a resource allocation process and a decision-making process. — *Interview 11*

Second, as the projects themselves became riskier and costlier, local decision-makers began to recognize the value of accurate forecasts in helping them minimize risk.

> All the easy projects have been done. I don't think there are any easy projects left out there. If they were easy, they would be done already. What you have left now are ones that are technically extremely challenging, very costly, or politically very challenging — or all three of the above. The more rigor you can put into the planning process … and help people use that information to make decisions, the better off people are. They're making very expensive decisions these days. Projects are in the half-billion-dollars or billions-of-dollars range. Those are very tough decisions to make when there are so many demands for limited pots of money, both locally and at the federal level. – Interview 11

A few interviewees from transit agencies likewise expressed the idea that the New Starts process created data and information that were useful to local decision-making. A transit agency executive credited his agency's success with their reliance on information from the New Starts process to guide internal planning.

> [Our success] had to go back to the fact that we did our homework. [Planning] was based more on market than on politics. We came up with the right management team and the right local funding sources first, and then applied for New Starts funding. Frankly, [on our very first project] we used [the project evaluation criteria] as our measuring stick for the value and effectiveness of the project. We didn't have any expertise back then. We didn't know what was a good project and what wasn't. So we actually ended up using the New Starts process in our local decision-making process to answer the questions: Is it a good project? Does it add value? What are its benefits? Those are the criteria that are embedded in the New Starts process. We have copied the essence of the New Starts process now in our MPO process for the evaluation of all of our long-range transit capacity improvement projects. I think it's served us really well. — *Interview 5*

A consultant explained that, since forecasts prepared for New Starts projects are subject to so much scrutiny, they have come to represent an industry standard for excellence.

> New Starts forecasts made with travel models have historically had the highest scrutiny. And they're the most rigorous types of forecasts that you can make, as far as reviewing the model and reviewing the reasonableness of the project

assumptions. With my staff, I always try to hold any forecast we do to that standard. In some cases it's not possible because the study doesn't require it. But generally we try to hold ourselves to that standard. We try not to make forecasts with bad models, to the extent we can. And we try to be really rigorous about the inputs, even when it's not needed. There are some differences [between forecasts prepared for New Starts projects and those prepared for other projects], but my approach has been, I try to make it always like the New Starts side. — *Interview 9*

An FTA staff member described how the introduction of the Summit software package represented a watershed in the ability to evaluate the underlying assumptions that were used in ridership forecasts. Although the purpose of Summit was to assist project sponsors in computing transportation system user benefits (a measure that included travel time savings for existing riders), it had the additional effect of providing greater transparency about a travel demand model's underlying assumptions.

An ancillary, but it turned out —in my view anyway— a more important result of the Summit software was that, for the very first time, it produced detailed reporting of the ridership forecasts. That was the equivalent of shining a light into a really dark box, and there was all sorts of pretty ugly stuff going on that you would normally have a very hard time finding because of the complex nature of ridership forecasting and all the things that can go wrong. It turns out that this software, which was designed to report the benefits of the project in detail, showed a whole bunch of stuff that was random error nonsense. There were all sorts of unintended things happening. And all of a sudden, the ridership stuff got a lot more rigorous. That produced its own reaction as well: Folks weren't happy that they had to fix all this stuff. … I don't pretend my perspective is the general perspective, but from my perspective, … when the project sponsors … ran the software and saw the results, we were all a whole lot better off in terms of seeing where there were mistakes happening that previously had gone undetected. So, the hope is that that helped clean up a lot of stuff that previously had gone without any attention. So the hope is that for projects since then, that ridership forecasts may be somewhat improved. — *Interview 8*

A consultant confirmed that, although the Summit program was no longer used after the requirement to report TSUB was discontinued, it had a permanent effect on the accuracy of ridership forecasts, since it allowed modelers to correct persistent errors in their models, and improved their knowledge and understanding of travel demand modeling.

Summit was a program FTA came out with in 2003. That was a game-changer. … 2007 was the first year that projects that had had their forecasts reviewed by the Summit program were constructed and put in operation, and there's a

> noticeable overall jump in accuracy from 2007 onward. That was a big game-changer because you could actually identify and describe the major problems that the model had. Previously, you would just be lost and swimming in too many numbers, and you couldn't actually figure out what the hell was going on except at a very deep level. Summit allowed you to look at it … and actually see that the model's not doing a very good job at all. It found that the model-development practice and model-application practice was pretty bad. And I would say bad to the point of being almost criminal or fraudulent. … That was a complete watershed moment. We unlearned more about what we knew than I had ever learned. I've learned more about forecasting in the last 13 years than I had in the previous 13, by a country mile. — *Interview 9*

An agency staff member described how travel demand models are constantly refined and improved, and suggested that this has led to improvements in forecast accuracy over time.

> The travel forecasting abilities have been progressing. They're getting more and more complex, but we're getting more sophisticated and hopefully better at forecasting how people are using the system and how people's travel patterns will change as a result of improvements. For instance, the park-and-ride model wasn't a good predictor of [park-and-ride] use … in the early 2000s. We've since done a lot of research into how people use park-and-ride, how they make the choice of parking, and the characteristics of the households they come from. I think we're a lot better now, and we no longer have these arbitrary park-and-ride sheds that we'd used previously. So we've sized the park-and-rides based on a more reasonable estimate of demand. — *Interview 3*

Another interviewee observed that, since travel demand models were originally developed to determine highway capacity needs, many refinements have been required in order to appropriately incorporate travel by various public transit modes, and that many travel demand models continue to poorly represent short trips and non-commute trips.

> Most travel demand models are highway-based models, so it takes an investment on the part of a metropolitan area or the MPO to capture information on boardings for transit and to look at where transit trips originate and where they end. … Over the last five to six years, with advent of streetcars, it's become more important to pick up within-zone trips and short trips, or the discretionary trip that isn't a work-based trip. Our current challenge is how to develop a model that is more sensitive to those induced trips, those discretionary trips, those off-model trips, that are not work-based. Those are trips that happen throughout the day or on weekends because the availability of transit. … Because of land use and economic development changes, I'm using the transit system differently than the way we've historically used the transit system. — *Interview 7*

In 2013, FTA introduced STOPS as a simplified travel demand model that project sponsors could choose to use instead of their own regional travel demand models. A consultant described how the introduction of STOPS has improved modeling practice by allowing consultants to build forecasts from the ground up from a model that was designed specifically for transit, rather trying to incorporate detailed transit information into regional models that were originally designed to represent travel by private vehicle.

> The STOPS model has really helped because it's gotten us out of the game of trying to fix local travel models, which are typically geared toward policy analysis or traffic design volume needs and almost have nothing to do with anything else. So it's really gotten us out of requiring large amounts of surgery needed to local models to fix what's wrong with them or to make them coherently explain transit patterns. Now we can use STOPS and we get around it completely. — *Interview 9*

Another consultant described how the use of the STOPS model can streamline the forecasting and model review process.

> The atmosphere in the industry seems to be that FTA would rather have us use STOPS, and then we don't have to do these complicated reports. [FTA staff] don't have to dig into the intricacies of the model to try to make sure that nothing funny is going on. They take the STOPS output, compare it against multiple projects, and do it that way. I don't think that that's necessarily a bad idea. — *Interview 13*

The same consultant suggested that the use of STOPS might free model developers to develop models that may not be well suited to modeling for New Starts projects, but that had other strengths that would still be useful in long-range planning for metropolitan planning organizations.

> I think there's a wider question that some of us have been wondering in the back of our minds, which is — this whole modeling system was built in the era of large, major infrastructure investments. … Is that kind of modeling still going to be useful in an era of maintenance and incremental expansion? And I don't think anyone has a good answer for that, … but when the questions people ask with [models] have changed, does the model structure need to change too? Perhaps it does. Could we build a new model that really leverages big data? Maybe it doesn't do all the things that our current model does, but maybe it does some different things better. Those are all big questions. If FTA decides, "We don't care about individual choice modeling; we want everyone to use [STOPS]," I think in some ways it would free up MPOs and state DOTs to consider, "What else could our current models do? What else can they be used

> for?" … Maybe states and cities can use their models for other things. — *Interview 13*

A consultant expressed hope that STOPS might represent the beginning of a movement toward using multiple models to represent various types of travel at various scales, rather than seeking to represent all travel in a region with a single, comprehensive regional model.

> The modeling industry —not the forecasting industry, but the modeling industry— really pours its heart and soul into one model per region. They try to make this one model with a huge crystal chandelier of wonderful intricate design. … They do it because it's only one model and one model is easier to maintain than multiple models, but as a forecaster, I don't understand why they're not in favor of having multiple models to look at different things. And I think that's something that we need to think about going forward and I think it's also one of the hindrances of progress in forecasting. Especially with models getting more and more complex, the data's not keeping up. The number of people in this country who can actually adequately maintain a model is going down. This is why STOPS is a big deal because I think we're on a crash course on the modeling side. Very few people understand how to use these complex models that probably weren't designed very well to begin with. So I think developing forecasts with multiple models is probably a saner way to go. — *Interview 9*

The suggestion that multiple simple models would perform better than a single, complex model is well-supported by prior research findings that complex forecasting models generally do not perform better than simple models, especially when input data are uncertain (Alonso 1968; Ascher 1979; Makridakis et al. 1982, 1993; Makridakis and Hibon 2000). These conclusions may argue against a movement toward activity-based models that are intended to better represent underlying behaviors that motivate all types of urban travel.

However, opinions about the STOPS model were mixed among interviewees. A transit agency staff member expressed pride in his region's travel demand model and described the STOPS model as the "lowest common denominator."

> [The FTA has] said "Alright, there are so many deviations between regional models, … we're going to try and make a lowest common denominator by taking some of those choices —with land use being the biggest one— out of the hands of the project proponents."… You can still choose not to use it. For [our] area, we have a pretty sophisticated regional model. We're able to use that. … We think our model is significantly more sophisticated and we can get a lot more information out of it. The STOPS model is not as granular. It can't tell you as much. It has to make a whole bunch of assumptions. So there's a risk

> that you miss some of the benefits of the project by just using the STOPS model. — *Interview 1*

A consultant suggested that the STOPS is so easy to use that it might allow local project sponsors to reduce their reliance on consultants for preparing forecasts.

> FTA has put out what's called the STOPS model, which is supposed to make it simpler for a forecast to be made and for FTA to agree to the forecast. While we do a lot of STOPS applications, I think that does make it more possible for project sponsors to do it themselves. And I think that's probably what's happening, for Small Starts in particular. — *Interview 4*

One interviewee suggested that reducing the reliance on consultants for ridership forecasting might encourage project sponsors to better incorporate forecasts into decision-making.

> I wish that transit agencies would rely on some of the aspects of New Starts more than they do. Right now they farm it out to the consultants and it's just a number. They get a paper and they throw it at FTA, and FTA looks at the paper. I think the agencies ought to use the data that comes out of New Starts more than they typically seem to do. And they ought to have more ownership than they seem to have. — *Interview 5*

If the STOPS model does reduce the reliance of project sponsors on consultants and thus enable project sponsors to take more ownership of forecast data, this change could improve performance-based decision making at the local level.

Another way in which an FTA policy may have improved planning practice is through the requirement to produce Before and After Studies. One transit agency staff member explained that the process of preparing a Before and After Study creates a helpful opportunity to reflect on the lessons learned from a project.

> So all of that stuff goes into the analysis for explaining why you were accurate or if you were not accurate, where do you think you could do better? So I think the Before and After Study, as long as it's not used to be a gotcha, is actually a pretty good process. — *Interview 1*

A consultant described how his or her firm had used the results of published *ex post* analyses to validate the results of a travel demand model:

> There were some questions as to who would ride [a commuter rail line]. Was the model giving a forecast that would be expected? As a response to that, I

> assembled post-commuter-rail-implementation surveys from ten or eleven different MPOs … [in] places that have implemented commuter rail and done an after survey, and I created a report of different ridership characteristics. … So we could compare what was coming out of [our] model to what has been observed in other cities, and try to say, "Okay, are these results meaningful or not?"— *Interview 13*

Another transit agency staff member echoed the sentiment that Before and After studies are useful, but suggested that many in the transit industry failed to take full advantage of the lessons that can be learned from other agencies' Before and After Studies.

> There are usually some recommendations at the end [of a Before and After Study] on how we can do things a little better. Even if we were pretty good and we were on target for the forecasts, what have we learned? One of the primary reasons for having a Before and After Study is not just for your own information, but for FTAs purposes in evaluating projects around the country, and then for other project sponsors to get a better understanding of some failings that they can learn from elsewhere. … I've personally read every single report that's ever gone out, and I've found them pretty useful for my purposes. I don't think what people have learned is shared enough around the country, but I've been trying to disseminate the findings, not just with others here at [this agency] but with our regional partners. I did a presentation at [a local university] … and that question came up of, "Are all these lessons learned really being used?" And I said, "Well, at [this agency] they definitely are, but I'm not sure about the rest of the country." I'd expect they're probably not, but it's a great set of information. I'm not sure that other public transit agencies are doing the same. They should be. — *Interview 3*

A consultant likewise expressed uncertainty about whether Before and After Studies are effective in helping to disseminate lessons learned from completed projects.

> The FTA summarizes them and sends them to Congress, so that's the main recipient, but the whole point of doing that is to come up with a better forecast in the future. That might be from that same agency saying, "Gee we blew this forecast; we don't want to do that again; how can we improve our models or input assumptions?" And I think it's useful for projects in other cities. So for example, as we do the Before and After study for [a project] that we're working on now, we're thinking, "How can we make this useful for other [similar] projects so that they don't make the same mistakes?" … I don't know that [other agencies do refer to completed studies]. I think that was the FTA's desire: that these would be made available and that they would know how to use them. I guess I don't know that that happens. — *Interview 4*

One consultant who had taken the time to study published Before and After Studies described how empowering it was to be able to estimate the likely accuracy of a forecast based on the performance of forecasts for similar projects.

> If somebody asked me the question, "How accurate are these forecasts?" or "How accurate have transit forecasts been over the last 5 or 10 years?" no one could actually say what those numbers were. I find that pretty abysmal that over 40 years we can't say that. … But now I can actually put a number on it. It's pretty empowering. And that's very helpful to tell clients that your ridership is 5,000, and that based off forecasts in the past five years, it'll be plus or minus 20 percent of that. … That's pretty damn powerful. That's a game-changer. — *Interview 9*

As forecasters better understand the level of accuracy that is associated with their forecasts, they are able to describe them as ranges rather than as point forecasts, which is consistent with recommendations by Pickrell (1992) and Murphy (1993). One interviewee likewise explained how precise point forecasts create a false sense of certainty that the forecaster generally knows is not warranted.

> Something comes out of a model or your spreadsheet and there's this idea of precision. Sometimes I want to say, "Well, do you expect me to know what color the cars are going to be, making the turn on the road?" It's a mathematical tool. It's not perfect. It's not defining reality in the future. It's a tool that's trying to generate information. … You can plot the numbers as a dot on a graph, but in fact, that dot is a big fuzzy blob because there's a range of accuracy around it. It's plus or minus something. I've often been put on the spot back in the early planning stages, with people saying, "Will you guarantee that price?" Or "Will you guarantee that ridership?" Well, no, I can't. … Something that the FTA has done well recently is when you produce a forecast —whether it's a cost estimate or a ridership forecast— is have a discussion about the uncertainty around that. Not as a cover-your-behind thing, but just that people understand what can change that forecast going forward? — *Interview 11*

Another interviewee explained how presenting ranges of forecasts is an opportunity to offer decision makers more complete information, and to some degree, to transfer the responsibility for decision making from the forecaster to the forecast users.

> I think a huge missed opportunity is to characterize forecasts of ridership and — also of costs— in terms of ranges. The tradition of pretending that we actually know exactly how it's going to turn out is … is preposterous, but that's been the tradition. … First of all, identifying uncertainties and trying to quantify their potential influence, and then … working to reduce the uncertainty, … would all make the whole discussion of these projects a lot smarter. It would place forecasters of costs and ridership in a less precarious position, where they look like they assume that they … can produce a single number, and risk looking foolish perhaps when the number turns out to be quite different for reasons that may be well beyond their control. … A wise travel forecaster … does everything possible to document all the different things that could happen

> that would make [actual ridership] turn out to be a number that is substantially bigger or substantially smaller [than predicted]. The analogy I use is the weather forecasters. Weather forecasters learned decades ago not to say that it will rain on Saturday, so you should not plan on going to the zoo. They say, well, it's a 30-percent chance of rain, and they leave it up to you to decide whether you should go to the zoo and risk getting rained on. — *Interview 8*

However, as an FTA staff member described, when project selection criteria are based on specific thresholds for project performance metrics, ranges of forecasts must ultimately be converted to point forecasts in order to apply the criteria.

> The ratings system has never really graduated to a view that can handle ranges of forecasts. The ratings system is populated with a bunch of measures that have break-points, so if you're above this break point, you get a better score than if you're below if, and if you're above the next break point, you get an even better score. That kind of mechanics, which are pretty useful in the kind of thing we try to do, does not lend itself to working with the complexities that you get into when you have a whole bunch of ranges of forecasts. So there is a collision there between the uncertainties that are in the forecast and the need that FTA has for something that's more crisp than just. Maybe it could be 30, or it could be 40. Well, if there's a break point at 35, you really want to know whether it's going to be 30 or it's going to be 40. So there's always been pushback to the notion that we should invite everybody to just give us ranges. The attempt at a compromise is to get the range, but also get the best guess, and we'll use the best guess in the ratings system. But the whole business of dealing with ranges has been more than the ratings system has been able to process. — *Interview 8*

A consultant also described reconciling forecasts presenting ranges with a ratings system that applies point thresholds to make funding recommendations, and explains that it takes more strength to report a range of values when the range includes the threshold value than to report a range that falls entirely above the threshold.

> I actually gave a range of ridership. And it was one of the first New Starts projects where we actually did a range of ridership instead of a point forecast. I think it was the first New Starts project that had a range where more than half of the range actually took it into a situation where if the FTA looked at it and they took the average of the range, we would not have qualified for New Starts funding. Previously, when we had reported forecasts as ranges, they were ranges of very high numbers, so it didn't take any strength to produce a range of numbers where all of them would give you a favorable result. … That was the first time in my life, and I think it's the first time in New Starts where I actually not only gave a range, but I also gave a range where most of the numbers were not favorable. — *Interview 9*

In recognizing the uncertainty that accompanies any forecast, an interviewee differentiated between errors that can be attributed to misunderstanding or misrepresenting the project itself, and those that can be attributed to external factors such as inflation or even changes in the political climate. In the case of external factors, it may be impossible to reduce the risk of unforeseen changes, and it becomes important to manage risk.

> Nobody did the math wrong. Nobody missed the quantity. The price of oil or the price of concrete went up due to external factors that they had no control over. Somebody with a crystal ball could say there's always a risk of that. There's a lot more of a risk management process now that FTA has you go through. A lot of people look at that as an onerous thing, but for my projects, we've actually expanded it beyond evaluating the risk of changes in cost and forecast and schedule. We also look at political and more local issues like community opposition, lawsuits, and change in political office. What happens if we get a new governor? That happened to me three times on a project: the first one supported it, the second one didn't, the third one did, and the current one is reluctantly supporting it. Those are risks to the project. What do you do about it? A risk management process around moving your project forward around the cost estimates, around the schedule, around the whole decision-making process is so valuable … and I applaud FTA for really being that embedded in the process. — *Interview 11*

Another interviewee worried that overestimating the risk of cost overruns could lead to inefficiencies when cities are required to set aside more money for a project than is really necessary, which could put a worthwhile project out of reach for a project sponsor that has the financial capacity to fund its share of a project's most likely costs, but not enough to fund its share of a worst-cast scenario.

> It's easier to throw money at the issue than it is to really look and say, "Okay, is our assessment of the risk really accurate, or were we overly conservative or overly worrisome about it?" It's an issue because … you tie up that money and you can't get access to it. The burden is really on the locals because you're also tying up local money. — *Interview 7*

The same interviewee described how conservatively low ridership estimates might lead a project sponsor to under-size the project, leading to costly retrofit projects that wouldn't have been necessary if adequate capacity had been included in the initial project.

> In [a particular city], we thought the ridership estimate that we took in to get the project a full-funding grant agreement was way too conservative. When the project opened, ridership was much closer to the city's estimate than the FTA

> estimates. [The overly conservative FTA estimates] led to decisions that now have to be rectified. We wanted to build three-car platforms, and now as we expand the system, we have to go back and retrofit all the platforms on the initial alignment to go to three cars. — *Interview 7*

Another interviewee suggested that this conflict might be avoided if separate forecasts could be used for capacity planning, for revenue planning, and for project evaluation, since each of those forecast uses requires a different level of accuracy and has different consequences for under- or overestimates.

> One of the problems that I've been noticing is that we use the same forecast for designing the capacity of the system as for project evaluation, which is really what New Starts is all about. And those same forecasts are also used to project revenue. Those three items require different levels of accuracy. For capacity planning, you clearly want to be optimistic because adding capacity to a transit system is expensive. But when you're doing revenue projections, you actually want it accurate. … I would argue that we probably need to start thinking about using different forecasts for different things. Like for revenue and for project evaluation, we should use reference class forecasts so we can get an accurate revenue figure. For system capacity, I think the forecasts coming out of the model are actually okay for capacity analysis. — *Interview 9*

This suggestion that the appropriate degree of accuracy and bias may depend on what a forecast will be useful is consistent with Murphy's (1993) discussion of forecast value or usefulness. Murphy argues that, if the forecaster does not know what decisions a forecast will be based on, the value of the forecast is optimized when its accuracy is minimized. However, in the case of cost and ridership forecasts for transit projects, the forecaster does have good information on which decisions the forecasts will be used to support, and it may be appropriate to maximize forecast value by creating different forecasts to support different types of decisions. Alternatively, if forecasts are prepared as ranges of possible values, it might be appropriate to use the high end of the range for capacity planning and the low end of the range for revenue forecasts, in order to be conservative in both cases.

Ultimately, when New Starts policies and project evaluation criteria are viewed as an aid to facilitating excellence in transit planning practice, the motivations of transit professionals at the local and federal levels and in the public and private sector are much more aligned than when federal policies are seen as a tool to limit the federal role in public transit or as an arbitrary set of rules in a competition. An interviewee who was nearing the end of a long career in in the transit industry expressed pride in the work that he had done over the course of his career and also expressed confidence that, as federal policy continues to evolve, it was moving in the right direction.

> [The FTA staff have] some tools that they use that I have some real heartburn over, but these are the early days of them, and down the road, they'll get that sorted out too. I'm sort of retired and I don't have to worry about it anymore. … I'm delighted that this is what I did for 35 years. I feel like I can look back and say I did good planning, and I did it right, and I can stand by everything I did. — *Interview 11*

Conclusion

The combined results of these interviews offer important insights into perceptions in the public transit industry of the reasons for and consequences of changes in federal transit policy on transit planning and forecasting over the years. They also suggest testable hypotheses about how forecasting and planning for fixed-guideway transit projects may have changed over the past half-century.

Based on comments from the interviews, we can divide the history of the New Starts program in terms of three distinct time periods, as Table 4 summarizes.

Table 6: Changes in perspective on forecasting over time

Approximate time period	Prevailing view of the purpose of forecasts
Before 1989	To limit the role of the federal government in funding capital transit projects
1990 - 2007	To achieve fair competition among grant applicants for scarce federal funds
2007 - present	To assist in local decision-making.

Before the end of the Reagan Administration in 1989, the perceived purpose of New Starts project evaluation was to limit the role of the federal government in funding local transit projects. Throughout the 1990s and continuing into the early 2000s, perceptions shifted toward an emphasis on fairness and competition among cities competing for scarce federal resources. The introduction of the SUMMIT software program marks a transition in perceived emphasis from fairness to improving local planning practice.

How might these changes in the perceived purpose of project evaluation have influenced the accuracy of cost estimates and ridership forecasts submitted as part of applications for New Starts funding? During the first period, when the New Starts program was perceived to be antagonistic toward any federal funding for New Starts projects, the motivations for submitting optimistic forecasts may have been greatest, since project sponsors may have perceived a need to overcome federal antagonism by presenting proposed projects in the best possible light. During the second period, when project sponsors focused on competitiveness with other cities and regions that may be submitting applications with optimistically biased forecasts, incentives may have shifted towards simple matching optimism bias with that of competing project sponsors in order to level the metaphorical playing field. Finally, in the current time period, the perception that New Starts policies are intended to improve local planning practice and support prudent local decision-making might serve to improve forecast accuracy by aligning the incentives of the federal and local government.

In the following chapters, I test for the effects of these changes in perception by examining changes in the transit planning practice and forecast accuracy of cost estimates and ridership forecasts over time. In Chapter 6, I test for statistical relationships between the year in which a project's initial forecast was prepared and the characteristics of the project, including physical characteristics, financial characteristics, and characteristics of the pre-existing transit system. In Chapter 7, I test for relationships between forecast accuracy and the time at which a forecast is prepared, as well as relationships between forecast accuracy and the project characteristics analyzed in Chapter 6.

Chapter 6. Changes in Practice over time

This chapter describes the degree to which the changes in perspective and transit planning practice described in the preceding chapter have been accompanied by measurable difference in metrics that describe various aspects of transit planning for New Starts projects. As described in Chapter 4, I selected nine metrics grouped into four categories to describe planning practice as it relates to each project:

- Duration
    - Project development duration
    - Project construction duration
    - Ridership forecast horizon
- Financial characteristics
    - Total project cost
    - Federal funding share
- Physical characteristics
    - Project mode
    - Project length
- Local experience
    - Project sequence (whether a project was the first of its kind locally)
    - System mileage share (the project's share of the resulting network)

The most substantial changes in planning practice over time have been in terms of projects' durations and physical characteristics. The elapsed time between preparing a forecast and observing the actual value has declined, since less time has been required to construct a project, although the time from forecast preparation to the beginning of construction has not changed. The time to the horizon year selected for ridership forecasts increased from the earliest years of the New Starts program until the 1990s before decreasing sharply thereafter. In terms of physical project characteristics, the average project length has not changed substantially over time, but there have been changes over time in the most common project mode. For example, the earliest New Starts projects were mostly heavy rail, while bus rapid transit projects have appeared much more recently. The following sections describe these temporal variations in further detail.

Duration

I have divided the time between forecast preparation and cost or ridership observation into three components: the time from forecast preparation to the beginning of construction (referred to here as the project development duration), the time it takes to construct the project, and (in the case of ridership forecasts) the time from project opening to the forecast horizon year (assuming an ideal case in which ridership is measured in the same year referred to in the forecast). Project development durations for New Starts projects have not changed substantially over time, although construction duration has declined steadily. The total forecast horizon for ridership forecasts (the difference between the year in which the forecast is prepared and the year for which the forecast was made) has varied over the history of the New Starts program, and peaked under ISTEA in the 1990s.

Project development

For purposes of this analysis, project development duration refers to the time between the preparation of the first forecast (either cost or ridership) for a project and the beginning of project construction. In all cases, this value understates the total time period during which project planning and development took place, since forecasts can only be prepared once initial decisions about routes and station locations have been made, and the time required to make these initial decisions might vary considerably. Moreover, I define the first forecast as the earliest cost or ridership forecast that approximately corresponds to the project that was ultimately constructed. In the case of the Washington DC Metro and the Los

Angeles Red Line, the scope of the project changed dramatically shortly before (or even after) construction began, requiring the preparation of new forecasts (Schrag 2008; Elkind 2014). Thus, the project development duration for these projects is defined as zero, although planning for each of these two projects took place over decades preceding the relevant forecasts.

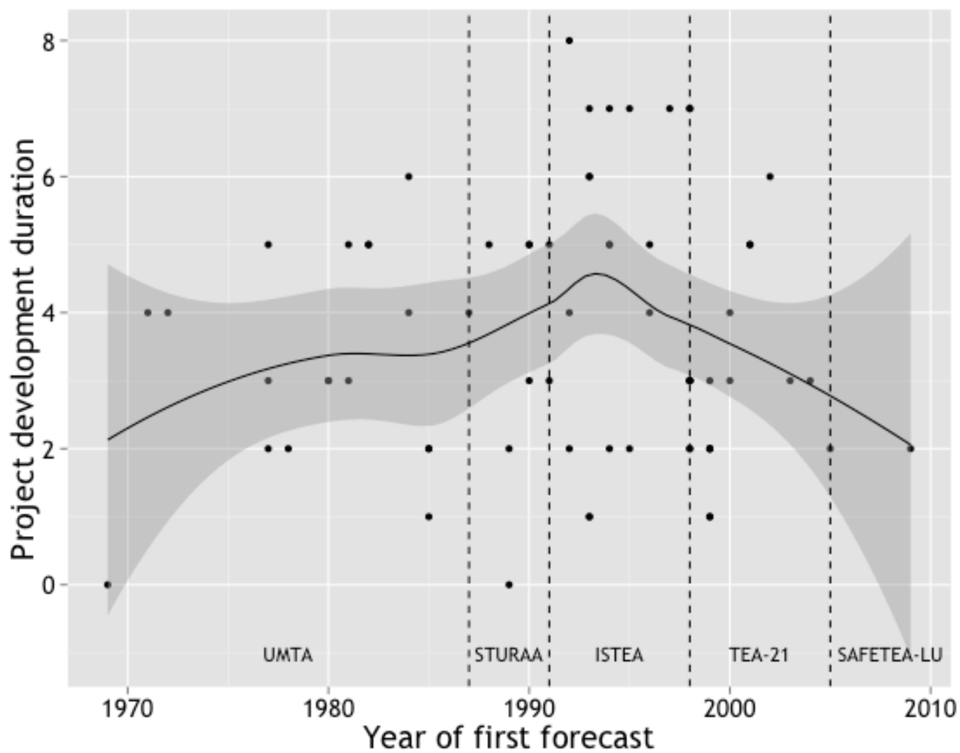

Figure 4: Variation in project development duration over time

Figure 4 illustrates the variation in project development duration over time, with a solid line for the locally weighted regression trend and darker shading for the 95-percent confidence interval around that trend. As shown, although project development duration has varied considerably, there is no clear trend over time. The 95-percent confidence interval for the correlation between the year of a project's first forecast and the subsequent project development duration is between -0.19 and 0.29. Since this interval includes zero, the correlation is not significant at a 95-percent confidence level.

I applied Tukey's Honestly Significant Difference (HSD) test to determine whether there were significant differences in project development duration associated with the New Starts authorizing legislation in place at the time of the forecast. The results are shown in Table 7 and indicate that there have not been significant differences in project development duration related to the current authorization authorizing New Starts grants.

Table 7: Differences in project development duration by current authorizing legislation

95-percent confidence interval for difference in average project development duration (years, row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-2.0 to 2.2	-2.7 to 0.6	-1.8 to 1.9	-4.1 to 6.7
STURAA			-3.3 to 0.9	-2.3 to 2.2	-4.4 to 6.8
ISTEA				-0.6 to 2.9	-3.0 to 7.8
TEA-21					-4.2 to 6.7
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Construction duration

In contrast to project development duration, construction duration has steadily decreased over time, as illustrated in Figure 5. The 95-percent confidence interval for the correlation between the year a project begins construction and the construction duration is between -0.69 and -0.35.

As Table 8 shows, this steady reduction in construction duration has, for the most part not been accompanied by discrete changes associated with new authorizing legislation. The time to construct projects for which construction began during the TEA-21 years was about 2 months to five years shorter than for projects that began construction under the Urban Mass Transportation Act (and this change cannot necessarily be attributed to differences in federal policy or practice). Otherwise, there are no significant differences in construction duration by authorizing legislation.

Figure 5: Variation in construction duration over time

Table 8: Differences in construction duration by current authorizing legislation

95-percent confidence interval for difference in average construction duration (years, row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-1.7 to 4.4	-1.1 to 3.6	0.2 to 5.4	-1.6 to 13.9
STURAA			-3.1 to 2.9	-1.8 to 4.6	-3.2 to 12.8
ISTEA				-1.0 to 4.1	-2.9 to 12.6
TEA-21					-4.5 to 11.2
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Forecast horizon

Figure 6 illustrates changes in ridership forecast horizons over time. As shown, the average forecast horizon increased steadily over the first three decades of the New Starts program, and has decreased dramatically since then. The more recent decreases have offset the previous increases such that the correlation for between the year a forecast was made and the time to the forecast horizon year is not significant, with a 95-percent confidence interval between -0.05 and 0.42.

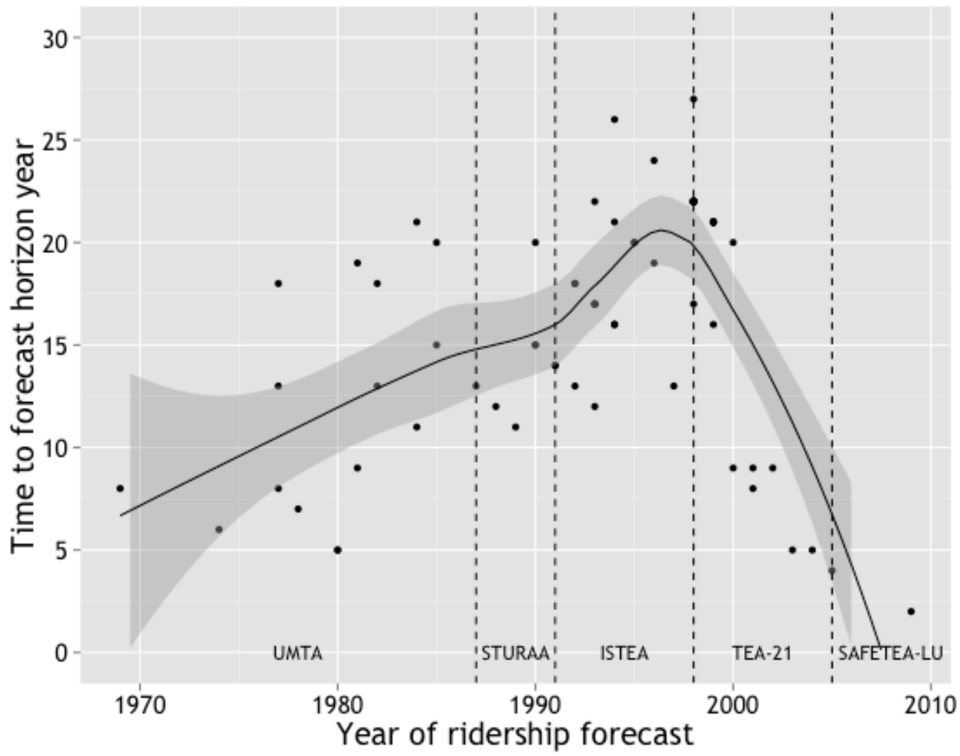

Figure 6: Variation in horizon duration over time

Table 9: Differences in forecast horizon by current authorizing legislation

95-percent confidence interval for difference in average forecast horizon (years, row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-8.1 to 4.5	-11.8 to -2.4	-7.4 to 3.0	-4.8 to 25.4
STURAA			-11.3 to 0.8	-6.8 to 6.1	-3.4 to 27.7
ISTEA				0.1 to 9.8	2.4 to 32.4
TEA-21					-2.7 to 27.6
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Figure 6 suggests that ridership forecast horizons peaked during the ISTEA years, and this is confirmed through the results of Tukey's HSD test that quantifies average differences in forecast horizons by authorizing legislation. As shown in Table 9, forecasts prepared during the ISTEA years had forecast horizons that were (at a 95-percent confidence level) two to twelve years longer than those prepared during the Urban Mass Transportation Act years, zero to ten years longer than those prepared during the TEA-21 years, and two to 32 years longer than those prepared during the SAFETEA-LU years.

The longer forecast horizons that were used under ISTEA might be a result of the increased role of metropolitan planning organizations (MPOs) in regional transportation planning, as ISTEA increased federal funding for MPOs and strengthened their role in selecting transportation projects (Federal Highway Administration 1995). Since MPOs are required to prepare long-range transportation plans covering a 20-year planning horizon, it would have been natural to apply this same horizon to any New Starts projects that were included in a long-range transportation plan. TEA-21 introduced the requirement for Before and After Studies comparing ridership forecasts to ridership observations two years after project opening. Such comparisons necessitated shorter-term ridership forecasts.

Financial characteristics

The financial characteristics of projects, as measured by total project cost and federal funding share, have been relatively consistent over time.

Total project cost

Figure 7 illustrates how the total cost of New Starts projects has changed over time. As shown, the first three projects in the sample (heavy rail projects in Baltimore, Washington DC, and Atlanta) all had particularly high total capital costs. Thereafter, the average cost of a New Starts project remained relatively constant. The 95-percent confidence interval for the correlation between the year of a project's first forecast and total project cost is between -0.43 and 0.02. Since this interval includes zero, the correlation is not significant at a 95-percent confidence level. Likewise, as shown in Table 10, there are no significant differences in total project cost by authorizing legislation.

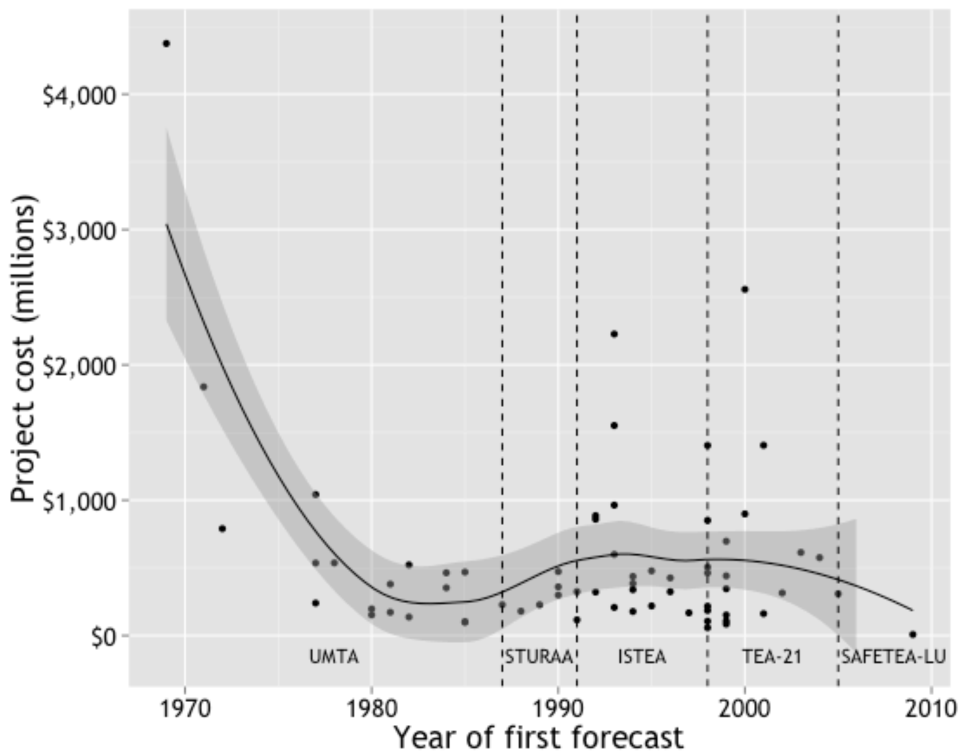

Figure 7: Variation in project cost over time

Table 10: Differences in project cost by current authorizing legislation

95-percent confidence interval for difference in average cost (billions, row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-\$1.1 to \$0.8	-\$0.6 to \$0.8	-\$0.8 to \$0.9	-\$1.8 to \$3.1
STURAA			-\$0.7 to \$1.2	-\$0.8 to \$1.2	-\$1.7 to \$3.3
ISTEA				-\$0.8 to \$0.6	-\$1.9 to \$3.0
TEA-21					-\$1.8 to \$3.1
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Federal funding share

As Figure 8 shows, federal funding shares for New Starts projects have varied dramatically, from as low as 20 percent (for the central and airport links of the Seattle light rail line) to as high as 100 percent (for the Medical Center light rail extension in Memphis, which was completed at 20 percent below budget, allowing its initial grant for 80 percent of anticipated project costs to cover the entire actual project cost). As described in Chapter 5, interviewees expressed a belief that federal funding shares had declined over time. However,

Figure 8 shows that such a decline is not apparent for the projects included in the study sample. The 95-percent confidence interval for the correlation between the year of a project's first forecast and the federal funding share is between -0.32 and 0.16. Since this interval includes zero, the correlation is not significant at a 95-percent confidence level.

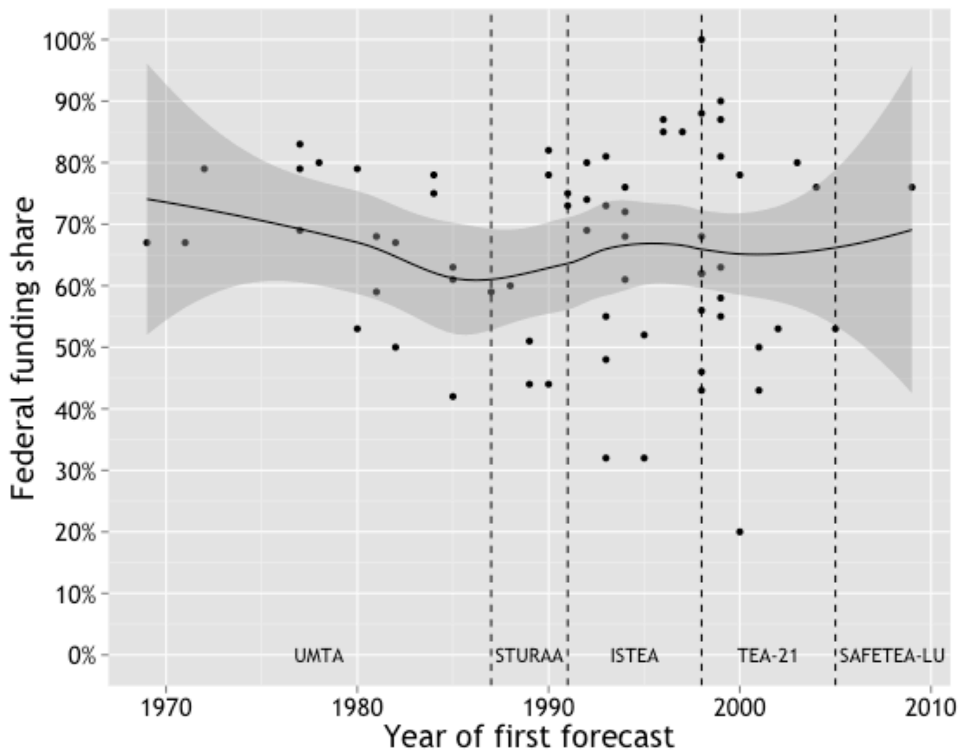

Figure 8: Variation in federal funding shares over time

Likewise, as shown in Table 11, there are not significant differences in federal funding shares by authorizing legislation. This is somewhat surprising in light of particular federal policies that would be expected to influence federal funding shares. For example, beginning with the passage of SAFETEA-LU in 1997, project sponsors have been permitted to keep a portion of the cost savings when a project is completed under budget (Duff et al. 2010). Since the federal funding share is calculated based on observed costs rather than planned costs, the degree to which a project is completed under budget would correspond with higher overall federal funding shares, all else being equal. Thus, this policy change might be expected to increase observed federal funding shares. In contrast, in 2003, the FTA explicitly changed its policy from a preference to fund New Starts projects with an 80 percent federal share to a preference for a 60 percent share (General Accounting Office 2003). This change would be expected to have generally reduced federal funding shares, on average. It is possible that the effects of these two policies, which took place within a few years of one another, served to counterbalance one another.

Table 11: Differences in federal funding share by current authorizing legislation

95-percent confidence interval for difference in average federal funding share (row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-16% to 21%	-14% to 15%	-11% to 21%	-56% to 38%
STURAA			-20% to 16%	-17% to 22%	-60% to 37%
ISTEA				-10% to 20%	-56% to 38%
TEA-21					-61% to 33%
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Physical characteristics

The physical characteristics of a project can be described in terms of both project length (in miles) and project mode. Although average project length has been relatively consistent over time, there have been temporal variations in the typical New Starts' project mode.

Project length

As Figure 9 shows, five projects in the sample have been longer than 25 miles. These were the initial heavy rail projects in Washington DC and Atlanta (the first two project forecasts in the sample) and three commuter rail projects during the TEA-21 years (a double-tracking project from Miami to Ft. Lauderdale, an initial line in Salt Lake City, and an initial line in Minneapolis). Outside of these five projects, average project length has been fairly constant over time. The 95-percent confidence interval for the correlation between the year of a project's first forecast and the distance covered by the project is between -0.30 and 0.17. Since this interval includes zero, the correlation is not significant at a 95-percent confidence level. Together with the finding that average total project cost has not changed substantially over time, this suggests that there has not been much change in the overall scale of projects receiving New Starts funding.

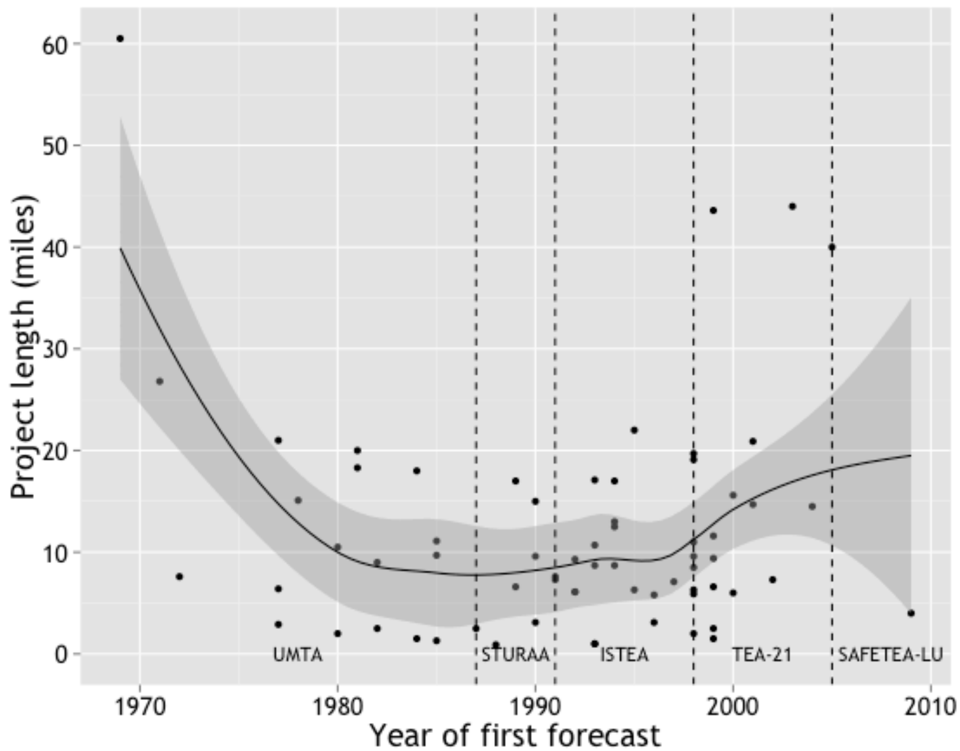

Figure 9: Variation in project length over time

Table 12: Differences in project length by current authorizing legislation

95-percent confidence interval for difference in average length (miles, row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-8.8 to 16.0	-5.7 to 13.3	-14.1 to 7.0	-22.5 to 40.4
STURAA			-11.9 - 12.2	-20.1 to 5.8	-26.9 to 37.7
ISTEA				-17.5 to 2.8	-26.1 to 36.5
TEA-21					-19.1 to 44.2
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Project mode

As discussed in Chapter 5, one interviewee suggested that changes in cost effectiveness metrics at various points in time have favored projects of particular modes.

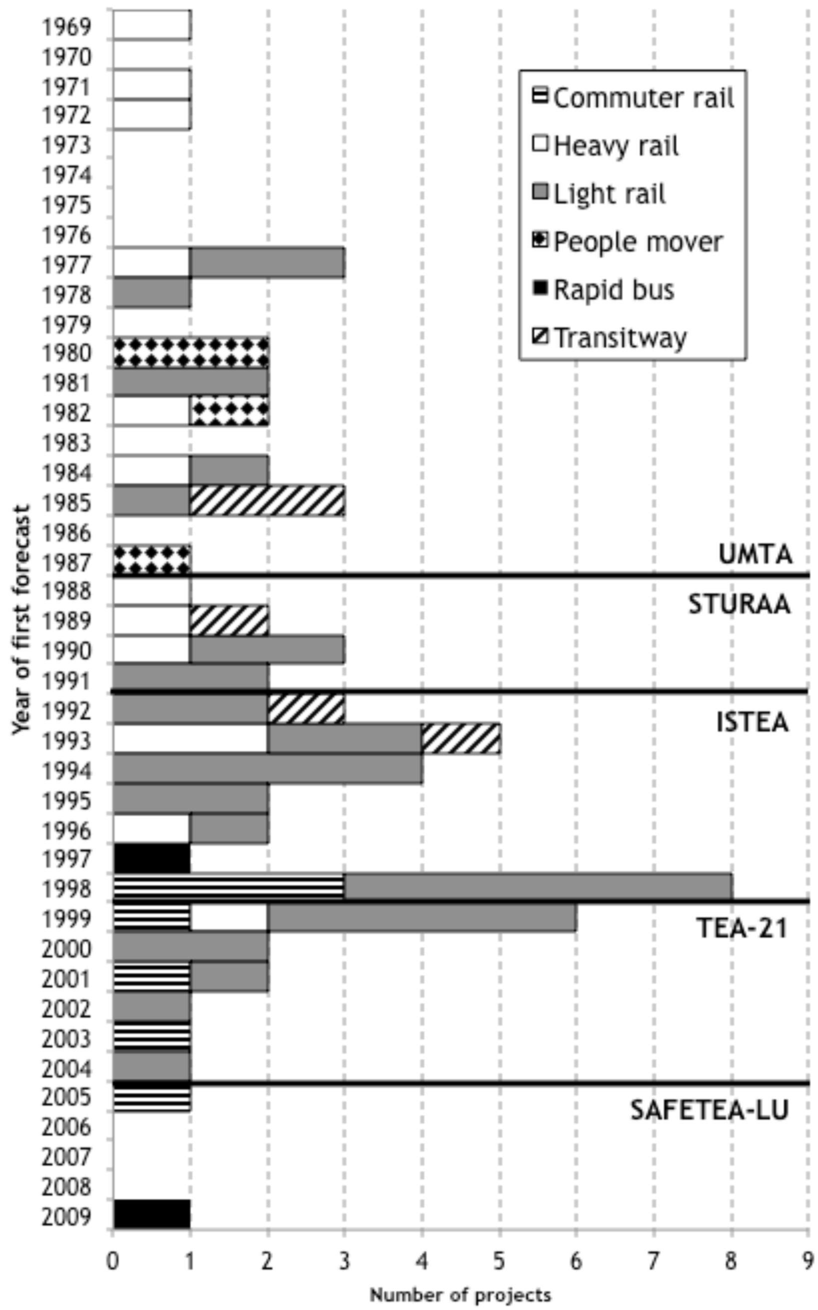

Figure 10: Project mode by year

Figure 10 shows the number of projects for each mode by year. It appears that particular modes have achieved peak popularity in subsequent approximate (and overlapping) time periods, as described below.

Heavy rail: 1969 – 1984

During the first fifteen years of the study period, initial forecasts were prepared for thirteen projects. Nearly half (six) of these were heavy rail projects, which represent half of all heavy rail projects in the sample. As described above, these initial heavy rail projects were also particularly large projects, both in terms of project length and total project cost.

People movers: 1979 – 1987

Between 1979 and 1984, initial forecasts were prepared for three of the four downtown people mover projects in the sample (in Detroit, Miami, and Jacksonville). The fourth, for which initial forecasts were prepared in 1987, was an extension of the Miami Downtown People Mover. These downtown people mover projects were the result of explicit efforts on the part of Congress to introduce new technologies to urban transportation.

In 1975, the Congressional Office of Technology Assessment issued a report opining,

> "it is unfortunate that there is no [automated guideway transit] installation in a city to ascertain feasibility. There should be a concerted effort by the Federal Government, municipalities, and the transportation industry to initiate a first urban application promptly" (Office of Technology Assessment 1975, 191).

In 1976, UMTA responded by calling for cities to submit proposals for downtown people mover projects (Merritt 1993). Sixty-eight cities submitted letters of interest, 38 of which ultimately submitted proposals (Merritt 1993; United States General Accounting Office 1980). Seven cities were selected for funding: Cleveland, Houston, Los Angeles, St. Paul, Baltimore, Miami, and Detroit (Merritt 1993; United States General Accounting Office 1980). In 1977, Congress requested that people mover projects in Jacksonville, St. Louis, and Indianapolis also be considered for funding, and UMTA awarded these cities (along with Norfolk, Virginia) technical study grants to further refine their proposals (Merritt 1993; United States General Accounting Office 1980).

By 1980, two of these eleven cities had abandoned their plans for downtown people movers: Cleveland because of the potential for negative impacts on the downtown area; and Houston because other transit modes were determined to be more cost effective (United States General Accounting Office 1980). In an audit of the nine remaining cities planning federally funded downtown people mover projects, the General Accounting Office (which has since been renamed the Government Accountability Office) found that the original purposes of the downtown people mover demonstration program were inadequate to justify continued funding for all nine projects that remained in the program and concluded that,

> "UMTA needs to justify why each of the presently planned [downtown people mover] projects is needed to meet program objectives and to show what each project will contribute to meeting them" (United States General Accounting Office 1980, 13).

Ultimately, only three of the original eleven projects were constructed: Miami, Jacksonville, and Detroit.

Transitways: 1985 – 1993

The study sample includes five transitway projects: the Seattle transit tunnel, and dedicated bus lanes in Houston, Boston, Denver, and Pittsburgh. Initial forecasts for all five projects took place in the eight-year period between 1985 and 1993. These projects can be viewed as precursors to the bus rapid transit projects that would gain popularity a decade later. As cost effectiveness standards came to be applied more stringently, transitways would have competed well with projects of other modes, especially people movers and heavy rail projects because they could be constructed at a much lower cost. However, the benefits of a transitway are more difficult to measure, since these projects do not necessarily represent new service and attributing ridership gains on existing bus routes to the construction of a transitway can prove to be complicated. The same can be said of any project that renovates an existing line rather than constructing a line for new service (such as the Douglas Branch reconstruction in Chicago, or double-tracking projects in Baltimore and Miami).

Light rail: 1990 – 1999

Just over half (36 out of 67) of the projects in the sample are light rail projects. Initial forecasts for two thirds of these were completed during the 1990s. This represents two thirds of all projects that had initial forecasts completed in the 1990s. These years of peak light rail construction loosely coincide with the ISTEA years.

Commuter rail: 1998 – 2005

Initial forecasts for all commuter rail projects in the sample were completed between 1998 and 2005. This time period loosely coincides with the TEA-21 years. During this period (in 2003), FTA began to require New Starts applicants to calculate Transportation System User Benefits (TSUB), which included travel time savings for existing transit riders, for inclusion in measures of cost effectiveness for proposed projects. Calculating benefits in terms of travel time savings would have been especially beneficial for projects serving long trips, such as commuter rail projects. Although initial planning for most commuter rail projects began before travel-time savings were explicitly considered in project evaluation, these projects benefitted from the TSUB metric by the time they were considered for advancement through later project milestones.

Bus rapid transit: 2006 – present

Only two bus rapid transit projects are included in sample: The Cleveland HealthLine and Mountain Link in Flagstaff Arizona. The HealthLine was initially identified as a baseline or transportation system management (TSM) alternative to a rail project that had been proposed for Cleveland's Euclid corridor. During the alternatives analysis for that project, the TSM alternative was selected as the local preferred alternative (LPA) (Federal Transit Administration 2012b). Thus, the only project in the sample that was initially planned to be a bus rapid transit project is the Mountain Link project in Flagstaff. This is also the only project in the sample for which initial forecasts were completed after 2005. In 2005, SAFETEA-LU created greater funding opportunities for BRT projects with the creation of the Small Starts program, which established a streamlined project for projects that are anticipated to cost less than $250 million and request less than $75 million in federal assistance (109th Congress 2005). All of the Small Starts projects in the sample are BRT projects. Streetcar projects have also qualified for Small Starts funds, but all of these have either not yet opened (such as the Seattle Streetcar City Center Connector) or have not been open for long enough to have completed a Before and After Study (such as the Tucson Streetcar).

The lack of rail projects in the sample with initial forecasts after 2005 can also be explained by the fact that bus projects can be completed more quickly than rail projects. The Mountain Link project opened for service just two years after initial forecasts were completed (Federal Transit Administration 2015b). Rail projects for which initial forecasts were completed after 2005 are either not yet complete, or have not been open long enough at the time of this writing to have completed Before and After studies.

Table 13: Time between planning of typical projects of each mode

95-percent confidence interval for difference in average year of first forecast (row minus column)						
	Heavy rail	**People mover**	**Transitway**	**Light rail**	**Commuter rail**	**Bus rapid transit**
Heavy rail		-9.7 to 14.9	-15.3 to 7.3	-15.2 to -1.3	-25.5 to -5.4	-34.5 to -1.8
People mover			-21.0 to 7.9	-22.1 to 0.5	-31.5 to -4.6	-39.4 to -2.1
Transitway				-14.5 to 6.0	-24.1 to 1.1	-32.2 to 3.8
Light rail					-16.1 to 1.6	-25.5 to 5.7
Commuter rail						-14.5 to 19.9
Bus rapid transit						

Gray text indicates that the difference is not significant at a 95-percent confidence level.

The order in which each mode has become most common can also be quantified by comparing the average number of years between the initial forecasts for projects of each mode (or the difference in the average year of the initial forecast for projects of each mode), as shown in Table 13. The initial forecast for an average heavy rail project was completed one to fifteen years earlier than that of an average light rail project, five to 25 years earlier than that of an average commuter rail project, and two to 35 years earlier than that of an average bus rapid transit project. The initial forecast for an average downtown people mover project was completed five to 32 years earlier than that of an average commuter rail project, and two to 39 years earlier than that of an average bus rapid transit project.

Cramer's V for the relationship between legislation and mode is 0.47, with a p-value (based on Pearson's Chi-squared test) of less than 0.01. This indicates that there is a significant relationship between authorizing legislation and project mode. I also tested the binary relationship between each possible combination of mode and legislation. The results are shown in Table 14.

Table 14: Relationship between authorizing legislation and project mode

Relationship between mode and authorizing legislation at the time of first forecast (Cramer's V/p-value)						
	Heavy rail	**People mover**	**Transitway**	**Light rail**	**Commuter rail**	**Bus rapid transit**
UMTA	0.19/0.21	0.40/0.01	0.07/0.93	0.21/0.14	0.21/0.19	0.11/0.91
STURAA	0.14/0.49	0.10/0.96	0.05/1.00	0.01/1.00	0.13/0.61	0.07/1.00
ISTEA	0.12/0.53	0.18/0.34	0.03/1.00	0.10/0.55	0.06/0.93	0.06/1.00
TEA-21	0.17/0.30	0.14/0.62	0.15/0.49	0.14/0.40	0.28/0.06	0.09/1.00
SAFETEA-LU	0.06/1.00	0.03/1.00	0.03/1.00	0.13/0.94	0.04/1.00	0.70/0.01
Gray text indicates that the relationship is not significant at a 95-percent confidence level.						

The only statistically significant relationships between pairs of modes and legislation are for downtown people movers, which are associated with the Urban Mass Transportation Act (initial forecasts for all downtown people mover projects in the sample were completed under the Urban Mass Transportation Act), and bus rapid transit projects, which are associated with SAFETEA-LU (the only project in the sample with an initial forecast completed under SAFETEA-LU is a BRT project).

Local experience

There does not appear to have been significant changes in the local, mode-specific experience of project sponsors over time, whether that experience is expressed as a categorical project sequence variable or as a continuous system mileage share variable.

Project sequence

As described in Chapter 5, some interviewees observed that, as the New Starts program matured, grant applications from veteran project sponsors who had already completed a project of the same mode as their proposed projects became more common.

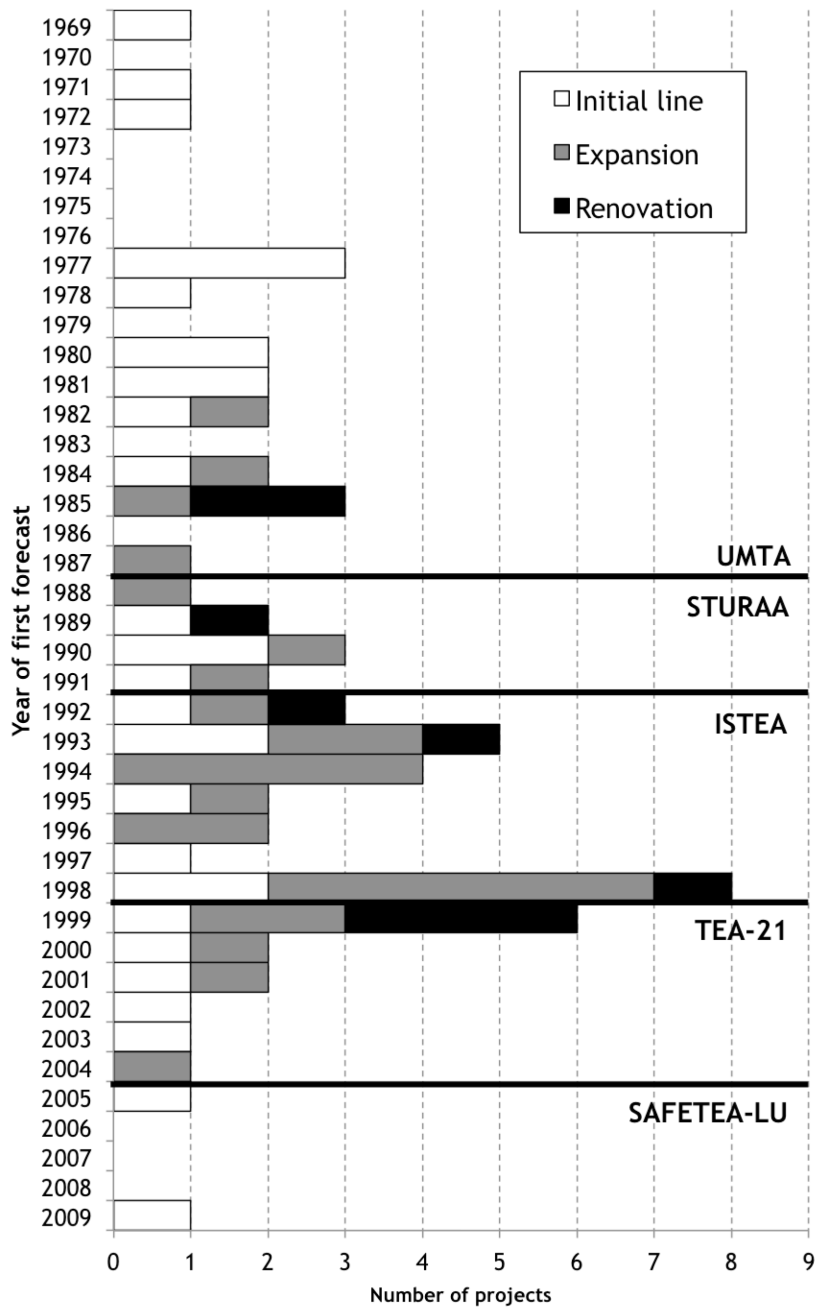

Figure 11: Project sequence by year

As illustrated in Figure 11, this is certainly true to some extent. The eleven forecasts in the sample, completed between 1969 and 1981, were all for projects that represented the initial line of a new fixed-guideway transit system. Thereafter, initial lines represented just 36 percent (20 projects) of the projects in the sample. Expansion lines represented 48 percent (27 projects), and renovation projects represented 16 percent (9 projects).

Table 15: Time between planning of typical initial, expansion, and renovation projects

95-percent confidence interval for difference in average year of first forecast			
	Initial line	Expansion	Renovation
Initial line		-10.7 to 0.1	-12.1 to 3.4
Expansion			-7.0 to 8.8
Renovation			
Gray text indicates that the difference is not significant at a 95-percent confidence level.			

Table 15 shows the 95-percent confidence intervals for the average number of years between initial forecasts for initial lines, expansion, and renovation projects. None of these differences are statistically significant.

Table 16: Relationship between authorizing legislation and project sequence

Relationship between mode and authorizing legislation at the time of first forecast (Cramer's V/p-value)			
	Initial line	Expansion	Renovation
UMTA	0.28/0.04	0.25/0.08	0.05/0.97
STURAA	0.01/1.00	0.03/1.00	0.03/1.00
ISTEA	0.29/0.03	0.30/0.03	0.01/1.00
TEA-21	>0.00/1.00	0.08/0.74	0.10/0.68
SAFETEA-LU	0.05/0.94	0.10/1.00	0.05/1.00
Gray text indicates that the difference is not significant at a 95-percent confidence level.			

Cramer's V for the relationship between authorizing legislation and project sequence is 0.28, with a corresponding p-value (based on Pearson's Chi-squared test) of 0.25. This indicates that there is not a significant overall relationship between authorizing legislation and project sequence. To better understand which legislative periods are associated with projects in each sequence category, I also tested the binary relationship between each possible combination of sequence and legislation. The results are shown in Table 16. Initial lines are associated with the Urban Mass Transportation Act, and expansion projects are associated with ISTEA. No other relationships are significant.

Share of system mileage

The three project sequence categories offer a useful way to think about local experience with projects of any particular mode. However, the expansion category masks large variations in the magnitude of an expansion. For example, it includes both the Colma BART station project, which was a one-mile extension of an existing 98-mile system, and the Denver Southeast corridor project, which was a 19-mile extension of an existing 5-mile system. This variation can be better described by a continuous variable representing the share of the resulting system the project represents (referred to in this research as a project's system mileage share).

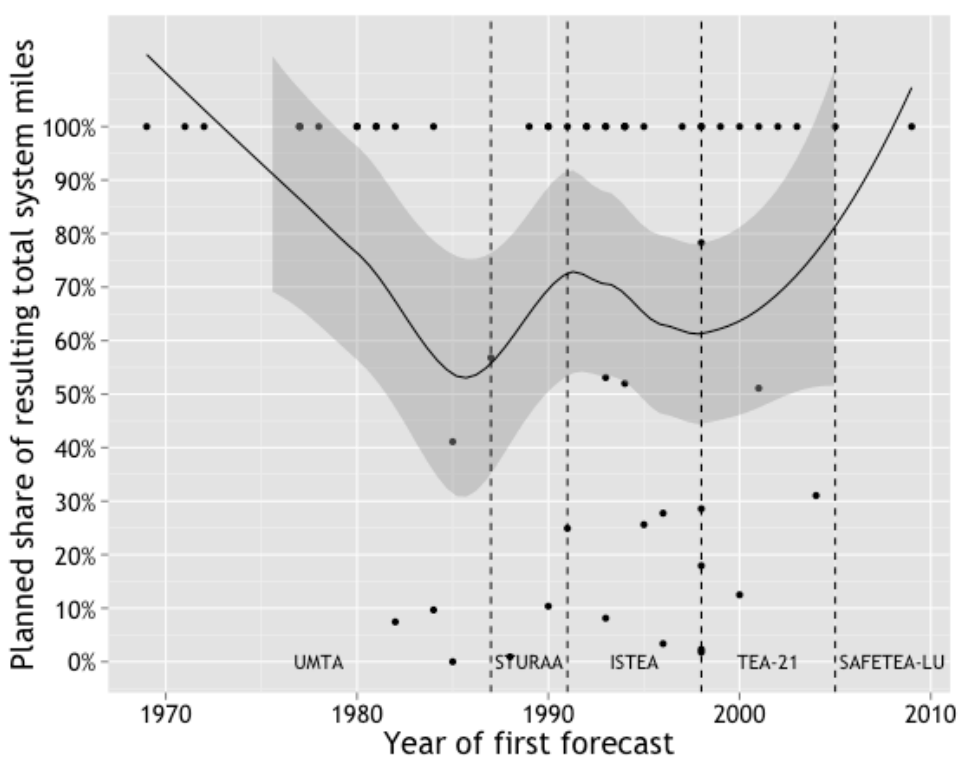

Figure 12: Variation in planned share of total system miles over time

As Figure 12 shows, sample projects' system mileage shares appear to have generally decreased during the Urban Mass Transportation Act years and ISTEA years, and increased during the STURAA and TEA-21 years. These reversals have tended to cancel one another out, such that there is no consistent trend in system mileage shares over time. The 95-percent confidence interval for the correlation between the year of a project's first forecast and the share of the planned network the project represents is between -0.42 and 0.09. Since this interval includes zero, the correlation is not significant at a 95-percent confidence level. Likewise, as shown in Table 17, there are no significant differences in average system mileage shares by authorizing legislation.

Table 17: Differences in share of system mileage by current authorizing legislation

95-percent confidence interval for difference in average share of system mileage (row minus column)					
	UMTA	**STURAA**	**ISTEA**	**TEA-21**	**SAFETEA-LU**
UMTA		-25% to 69%	-15% to 57%	-40% to 47%	-129% to 95%
STURAA			-47% to 44%	-70% to 33%	-154% to 76%
ISTEA				-59% to 25%	-149% to 74%
TEA-21					-135% to 93%
SAFETEA-LU					
Gray text indicates that the difference is not significant at a 95-percent confidence level.					

Conclusion

In general, I find that some of the changes in practice alluded to in Chapter 5 can be quantified and confirmed empirically. There are indeed distinct periods of time in which projects of a particular mode became more or less common, and there was a steady increase in the time until ridership forecast horizons until the 1990s (the ISTEA years), followed by a sharp decline thereafter. On the other hand, the quantitative analysis failed to confirm other observations made by interviewees. For example, initial lines of new projects have continued to represent a large share of New Starts projects, and perhaps most surprisingly, federal funding shares have been relatively consistent over time. Failure to observe these trends quantitatively does not necessarily mean that interviewees were mistaken. It is possible that the effects they describe have happened too recently to be included in the dataset, that the magnitude of the effects has been to small to achieve statistical significance, or that they are hidden by selection bias within the sample, which includes only those projects that were selected for funding. For instance, greater scarcity in federal funds may not be reflected in a smaller share of proposed projects receiving funding, rather than lower federal funding shares for those that do receive grants. However, the study sample only includes those projects that were ultimately constructed. In addition to the changes in practice that were mentioned by interviewees in Chapter 5, another significant change in practice has been a trend toward shorter project construction durations. How have these changes influenced forecast accuracy? That is the topic of Chapter 7.

The changes in practice described in this chapter, together with the changes in perspective documented in Chapter 5, all have the potential to influence the accuracy of cost and ridership forecasts. In Chapter 7, I analyze the relationships between each of these variables and the accuracy of ridership forecasts and initial and final cost estimates. I also use a series of regression models to determine the degree to which each of these changes independently contributes to observed changes in forecast accuracy over time.

Chapter 7. Correlates with Forecast Accuracy

As described in Chapter 4, I hypothesize that changes to the New Starts program over time have led to changes in transit planning and forecasting practice, and that these changes in practice have in turn led to measurable changes in forecast accuracy. Chapters 5 and 6 describe changes in planning and forecasting practice I identified through both quantitative and qualitative analyses. How do these changes relate to observed differences in forecast accuracy? That is the subject of this chapter.

I first describe how the variables discussed in Chapters 4 and 6 relate to the accuracy of ridership forecasts, initial cost estimates, and final cost estimates. Then, I describe the independent relationship of project characteristics and timing to forecast accuracy through a series of regression models.

Without controlling for other factors, six factors are most closely associated with more accurate forecasts: year of forecast preparation, construction duration, total project cost, federal funding share, project mode, and local unemployment rate. For both initial cost estimates and ridership forecasts, later years of forecast preparation and shorter construction durations are associated with greater accuracy. Lower-cost projects with higher federal funding shares are associated with more accurate initial and final cost estimates. Finally, forecasts prepared during times of low local unemployment and for light rail projects tend to be the most accurate for both final cost estimates and ridership forecasts.

The results of the regression models indicate that four of these variables have an independent relationship with forecast accuracy, even controlling for other factors. Project mode is associated with differences in forecast accuracy for all three forecast types (ridership, initial cost, and final cost). Additionally, shorter construction durations and later forecast preparation years are associated with more accurate ridership forecasts, and lower-cost projects are associated with more accurate initial and final cost estimates.

Correlations among errors

Pickrell (1989) and Flyvbjerg et al. (2005) have offered political explanations for inaccuracy in cost and ridership forecasts, suggesting that there are strong incentives for project sponsors to prepare optimistically-biased forecasts. Likewise, the interview results presented in Chapter 5 suggest that incentives for forecast optimism may have diminished overtime, which may explain some of the observed improvement in forecast accuracy described in Chapter 6. These political incentives might explain optimism bias in both cost and ridership forecasts. To the extent that they do apply to both, forecast error would be correlated with cost estimate error if the incentives to understate cost for a particular project would also incentivize project sponsors to overstate ridership. On the other hand, project sponsors might respond to political incentives for optimism by producing forecasts that are optimistic in terms of either cost or ridership, but not both (which could result in a negative correlation between cost and ridership error). However, my analysis of more recent forecast data suggests that neither of these two explanations appears to govern the observed relationship between ridership forecast error and cost estimate error, for either initial or final cost estimates.

As illustrated in Figure 13, it is certainly true that both cost and ridership forecasts have been optimistically biased for most projects. Of the 50 projects for which accuracy data is available for all three forecast types (ridership forecasts, initial cost estimates, and final cost estimates), a majority were optimistic in terms of both ridership forecasts and initial cost estimates, and a plurality were optimistic in terms of all three forecast types. However, as the two left panels of Figure 14 show, there is not a strong relationship between the magnitudes of ridership forecast errors and those of the errors for either initial or final cost estimates. The 95-percent confidence interval for the correlation between ridership forecast error and cost estimate error is between -0.37 and 0.13 for initial cost estimates and between -0.38 and 0.15 for final cost estimates.

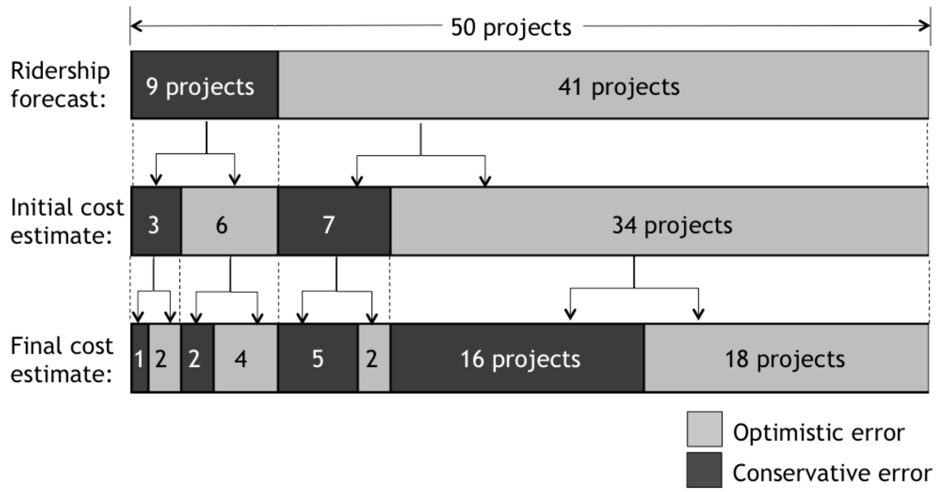

Figure 13. Presence of optimistic error by forecast type

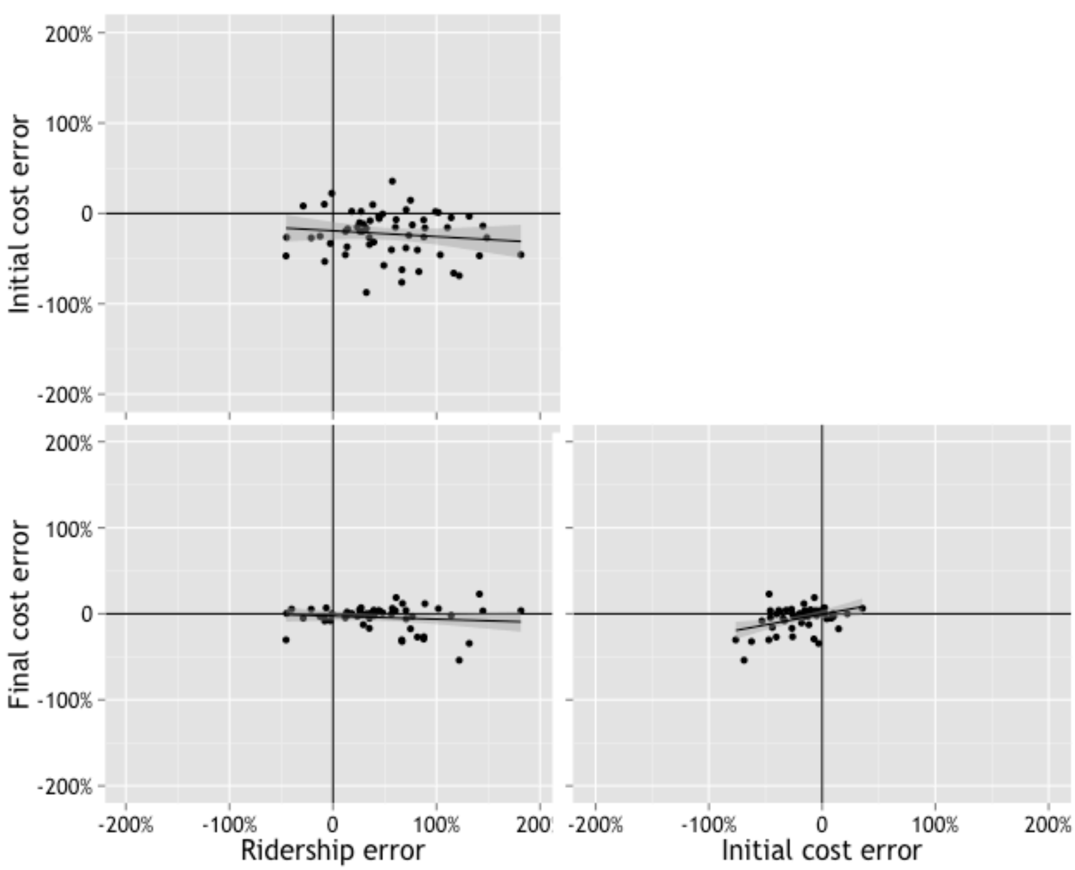

Figure 14. Relationships among errors in ridership, initial cost, and final cost forecasts

Figure 14 also shows the relationship between the errors of initial and final cost estimates in the bottom right panel. Errors for initial and final cost estimates are positively correlated, with a 95-percent confidence interval between 0.14 and 0.60.

Variation over time

Figure 15 shows the variation in the errors of ridership forecasts, initial cost estimates, and final cost estimates over time. Both initial cost estimates and ridership forecasts have improved in accuracy over time, while the error of final cost estimates has consistently hovered around zero (with a median of -0.6 percent).

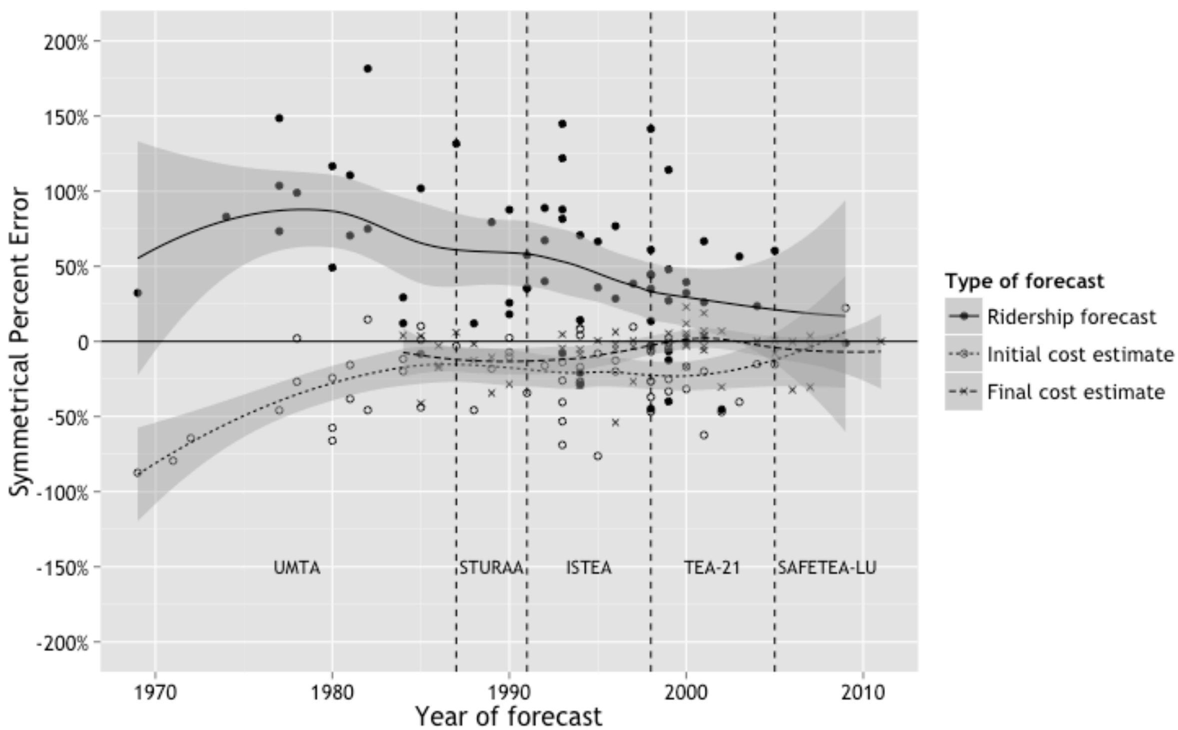

Figure 15. Variation in forecast error over time

These trends can be quantified by the correlation between the year a forecast was made and its error. Ridership forecast error is correlated with the year of the forecast with a 95-percent confidence interval between -0.59 and -0.18. Initial cost estimate error is correlated with the year of the forecast with a 95-percent confidence interval between 0.08 and 0.53. The correlation between the year of the final cost estimate and cost estimate accuracy is between -0.05 and 0.44. Since this interval includes zero, the correlation is not statistically significant. Thus, the accuracy of final cost estimates —which have generally been relatively unbiased— has not changed much over time.

I also used Tukey's HSD test to check for differences in accuracy in relation to the authorizing legislation in place at the time a forecast was made, and there were no statistically significant differences by legislation era.

Variation with project duration

Intuition would suggest that longer forecast horizons (the time between forecast preparation and observation of an actual value) would be associated with greater forecast error, because conditions generally are more certain nearer than further in the future; this suggestion has been confirmed in broader studies of population forecasts (White 1954; Keyfitz 1981; S. K. Smith 1987; S. K. Smith and Sincich 1992).

For capital cost estimates, the forecast horizon is the sum of the remaining project development time (from the time the forecast is prepared until construction begins) and the construction time. Cost estimate error can thus be divided into components corresponding to these two time periods: cost escalation refers to the increases in cost estimates over the course of project development; and cost overruns refer to increases in cost incurred over the course of construction (Love et al. 2013). Since ridership forecasts do not necessarily reference the project's opening year, the horizon for ridership forecasts is typically farther into the future than the sum of the project development duration and the construction duration.

As described below, the length of time before beginning construction is not related to the accuracy of ridership forecasts or initial cost estimates, but is related to the accuracy of final cost estimates. Conversely, construction duration relates to the accuracy of ridership forecasts and initial cost estimates, but not to the accuracy of final cost estimates. Finally, total forecast horizon is not correlated with ridership forecast error.

Remaining project development time

As shown below in Figure 16, there is minimal variation in accuracy associated with differences in remaining project development duration for either ridership forecasts or initial cost estimates. The 95-percent confidence interval for the correlation is between -0.22 and 0.28 for ridership forecasts and between -0.25 and 0.25 for initial cost forecasts. However, the magnitude of errors for final cost estimates does increase as the time until construction increases from zero to three years, with a 95-percent confidence interval for the correlation between -0.58 and -0.14.

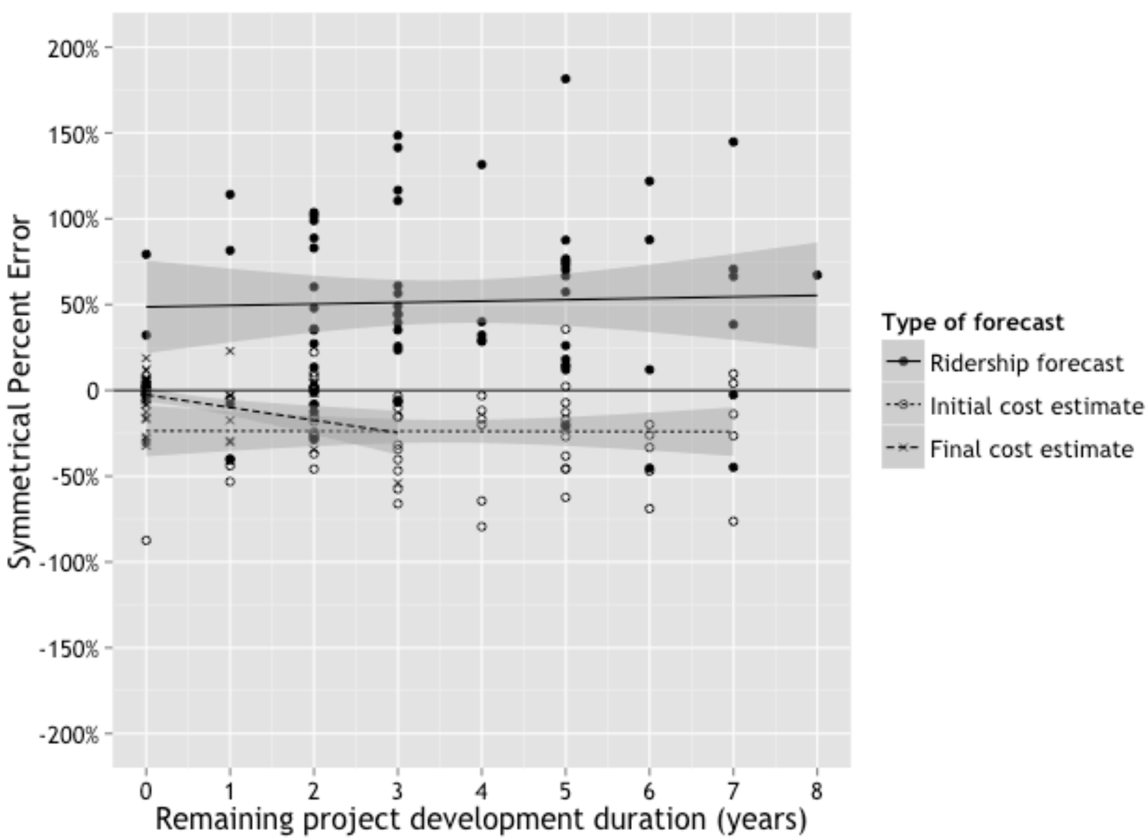

Figure 16. Forecast accuracy variation by remaining project development duration

Final cost estimates are typically prepared immediately prior to the beginning of construction. Among the 59 projects for which final cost estimates are available, three-quarters (44 projects) began construction in the same year the final cost estimate was prepared, and 86 percent (51 projects) had begun construction by the year after final forecast preparation. Thus, delays in the beginning of construction until two and three years after final cost estimates are prepared presumably represent unexpected circumstances that have affected both cost and schedule.

Construction duration

The relationship between forecast accuracy and construction duration is illustrated in Figure 17.

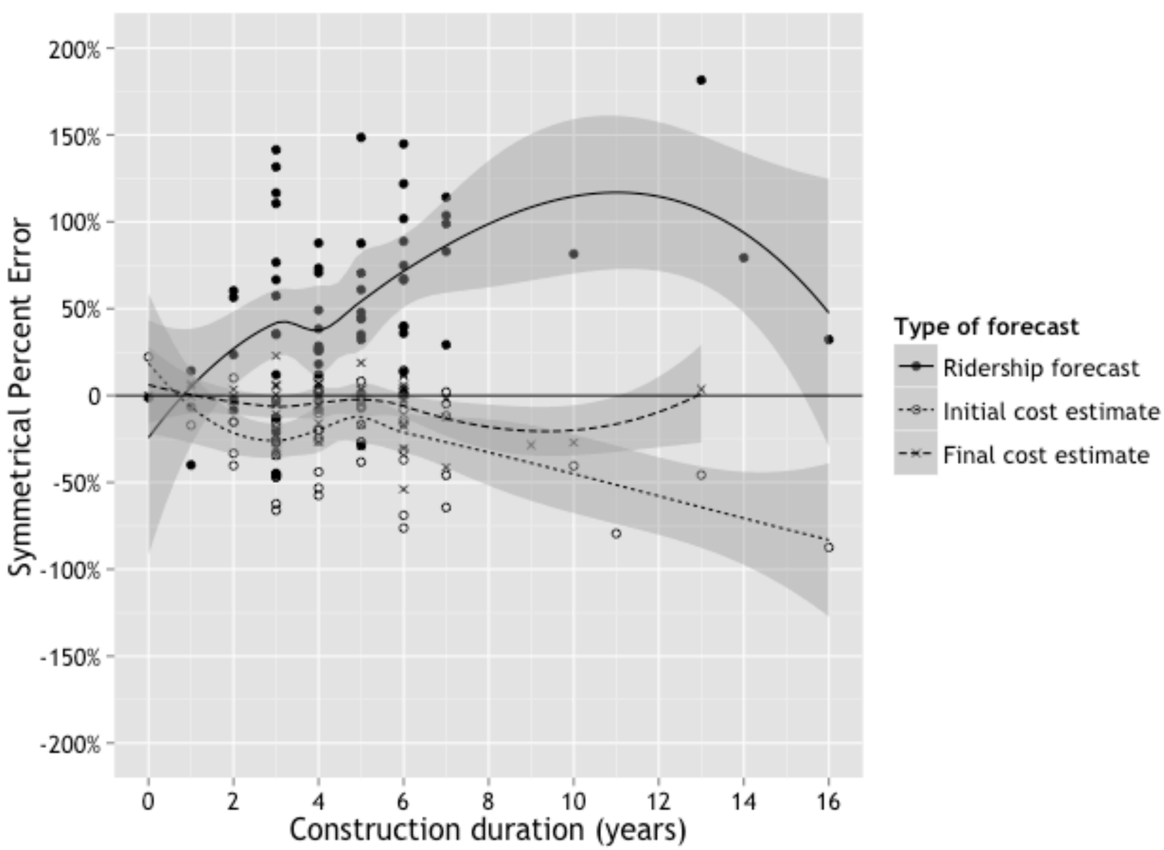

Figure 17. Variation in forecast accuracy by construction duration

In contrast to the relationships with remaining project development duration, accuracy correlations with construction duration are significant for ridership forecasts and initial cost forecasts, but not for final cost forecasts. The 95-percent confidence intervals for the correlations are between 0.16 and 0.58 for ridership forecast error and between -0.58 and -0.15 for initial cost estimate error. The 95-percent confidence interval for the correlation between construction duration and final cost estimate error is between -0.43 and 0.07.

Forecast horizon

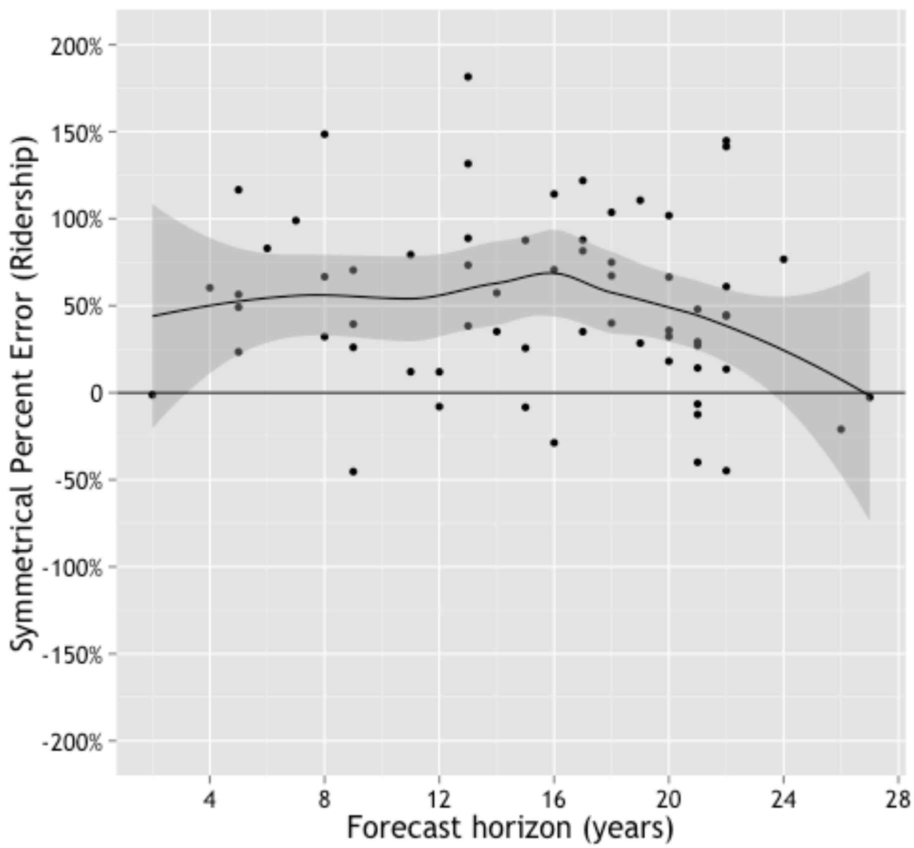

Figure 18. Variation in ridership forecast error by forecast horizon

Since ridership forecasts often refer to a year other than the project's opening year, the forecast horizon is later than the sum of the project development duration and the construction duration. Surprisingly, ridership forecast error is not correlated with forecast horizon, as illustrated in Figure 18. The 95-percent confidence interval for the correlation is between -0.37 and 0.11.

In many cases, ridership was not observed in the forecast horizon year, either because the project was not completed before the forecast horizon year, or because the observation was required for reporting purposes before the forecast horizon year arrived. For purposes of this analysis, when more than one ridership observation was available, the observation taken closest to the forecast horizon year was used.

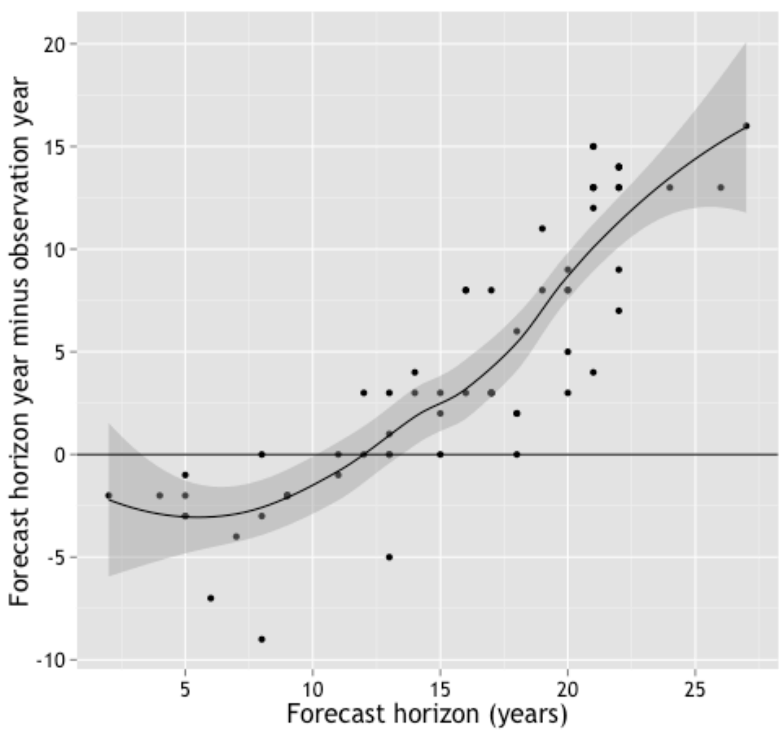

Figure 19. Horizon duration and time from horizon to observation year

Of the 64 ridership forecasts in the sample, only seven have ridership observations from the forecast horizon year. Over half of the projects in the sample (37 forecasts, or 58 percent) have observations within five years of the forecast horizon year, and over three quarters (49 forecasts, or 77 percent) have observations within ten years of the forecast horizon year. Sixteen have observations that are earlier than the forecast horizon year, and 41 have observations that are later than the forecast horizon year.

Since the case in which a ridership observation was taken before the forecast horizon year is more common for later forecast horizons, there is a strong correlation between forecast horizon and the difference between the forecast horizon year and observation year, with a 95-percent confidence interval between 0.79 and 0.92. This relationship is illustrated in Figure 19.

Variation with financial characteristics

Two financial characteristics of a project might be anticipated to relate to forecast accuracy: total project cost and federal funding share. Some studies on optimism in cost and patronage forecasts have focused on “megaprojects” (Flyvbjerg, Bruzelius, and Rothengatter 2003; Altshuler and Luberoff 2003; De Bruijn and Leijten 2007; Plotch 2015), which may suggest that very large, costly projects are particularly vulnerable to optimism bias. While this relationship is not observed for ridership forecasts, it is observed for both initial and final cost estimates. Additionally, Pickrell (1992) has suggested that reliance on federal funds may incentivize optimism, particularly for cost estimates. This explanation might be extended to suggest that higher federal funding shares would increase optimism bias. However, the data included in this study do not necessarily support this conclusion.

Total project cost

Figure 20 illustrates the relationship between forecast accuracy and total project cost. Since the distribution of capital costs for projects in the sample more closely approximates a lognormal distribution than a normal distribution, the x-axis in Figure 20 uses a log scale. Total project cost is not correlated with ridership forecast error (the 95-percent confidence interval for the correlation is between -0.23 and 0.26), but it is correlated with the errors of both initial (95-percent confidence interval between -0.64 and -0.24) and final cost estimates (95-percent confidence interval between -0.57 and -0.13). When total cost is log-transformed, the directions and significance of the correlations are unchanged.

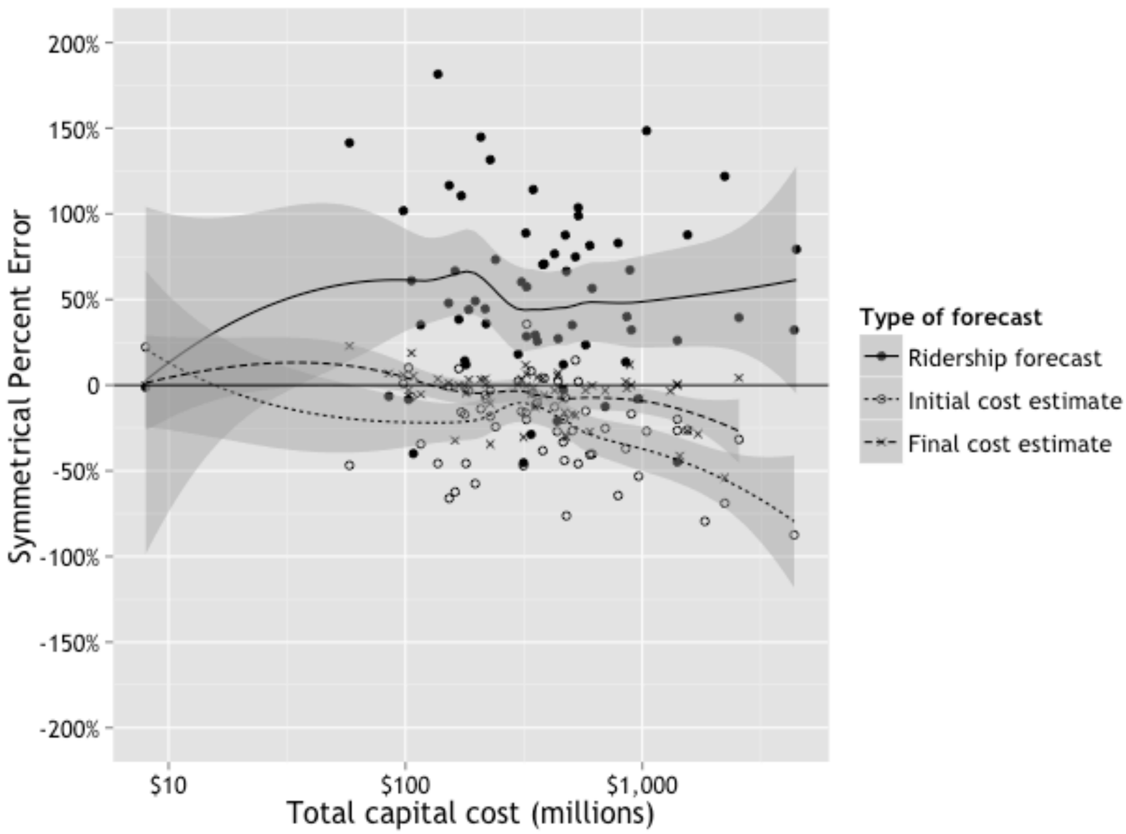

Figure 20. Variation in forecast error by total project cost

Federal funding share

The correlations between federal funding shares and forecast accuracy follow the same pattern as those between total project cost and forecast accuracy in the sense that they are significant for both initial and final cost estimates, but not for ridership forecasts. The 95-percent confidence interval for the correlation with ridership forecast error is between -0.25 and 0.24; for initial cost estimates, it is between 0.04 and 0.50; and for final cost estimates, it is between 0.30 and 0.68. These relationships are illustrated in Figure 21.

The correlations between cost estimate error and federal funding share are in the opposite of the expected direction — where the expectation is based on the hypothesis that higher federal funding shares incentivize more optimistic cost forecasts. This finding is due to the direct effect of cost estimate bias on actual federal funding shares, since grants are generally awarded to cover a specific percentage of capital costs, based on final cost estimates. Thus, cost overruns may reduce the actual federal funding share to a value that is less than the intended share. Likewise, completing a project under budget may increase the actual federal funding share to exceed the intended share. For this reason, federal funding share is omitted from the regression analyses predicting cost estimate accuracy presented later in this chapter, to avoid introducing endogeneity into those models.

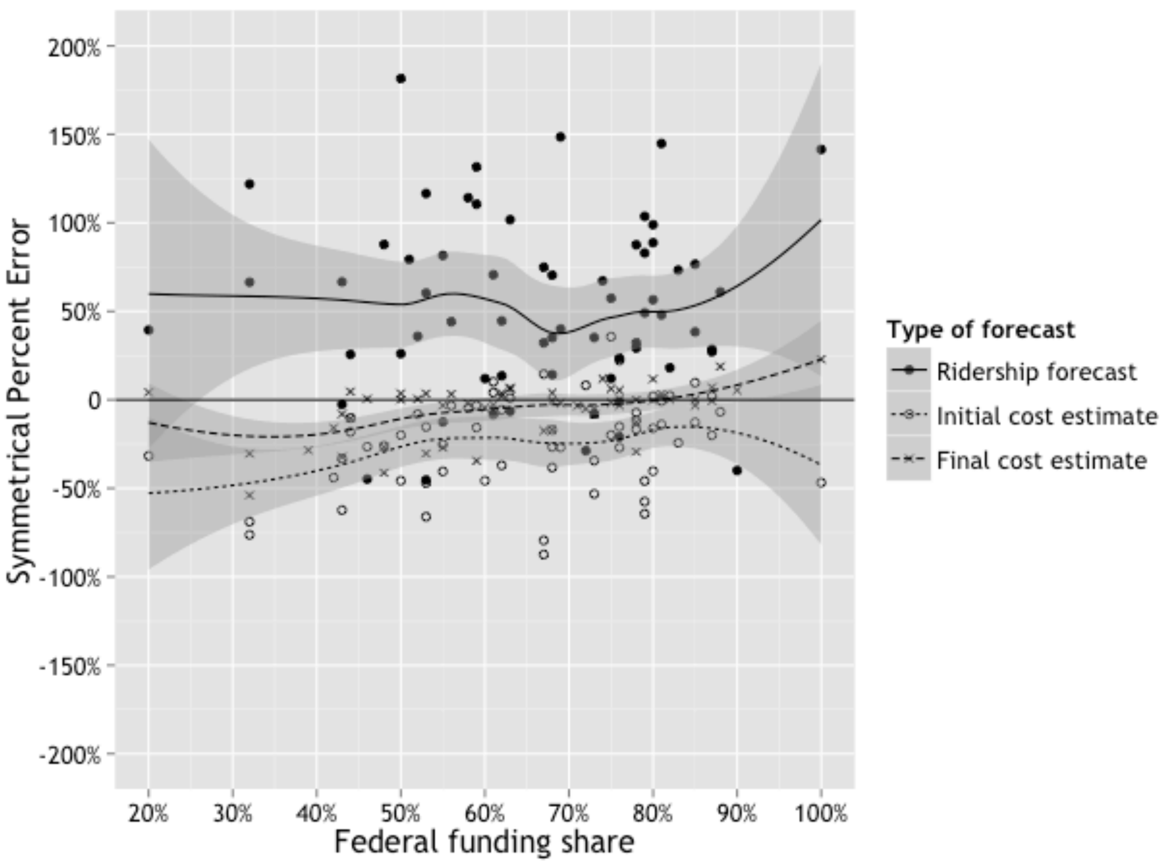

Figure 21. Variation in forecast accuracy by federal funding share

A better test for the hypothesis that high federal funding shares reduce incentives for forecast accuracy might be to use anticipated federal funding shares rather than actual federal funding shares; however, these data are not available for many projects in the sample, partly because it is not always clear whether project sponsors expect potential cost overruns or cost escalation to be mitigated by additional federal funding.

Variation with physical project characteristics

Neither of the two physical project characteristics included in the study —project length nor project mode— has a strong relationship with forecast accuracy.

Project length

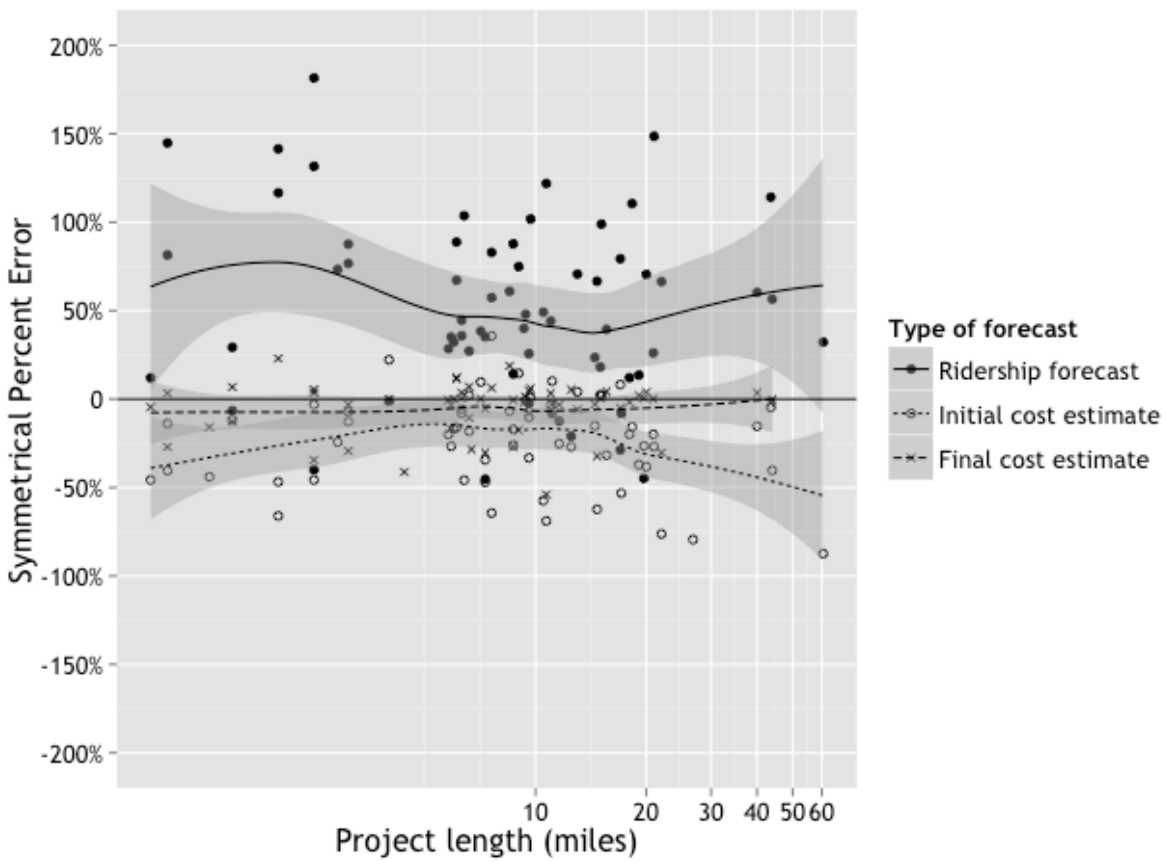

Figure 22. Variation in forecast accuracy by project length

The megaprojects hypothesis that large, costly projects are particularly vulnerable to optimism bias can be evaluated in terms of project length (in miles) as well as total project cost. The relationship between project length and forecast accuracy is illustrated in Figure 22. Like project costs, project lengths follow an approximately lognormal distribution, so the x-axis in Figure 22 is on a log scale. As shown, there is no clear positive or negative relationship between project length and the accuracy of any of the three forecast types evaluated.

None of the three forecast types have errors that are significantly correlated with project length. The 95-percent confidence interval for the correlation between ridership forecast error and project length is between -0.29 and 0.20; for the correlation with initial cost estimate error, it is between -0.47 and a positive value less than 0.01; and for the correlation with final cost estimate error, it is between -0.17 and 0.34. Correlations remain insignificant when project length is log-transformed.

Project mode

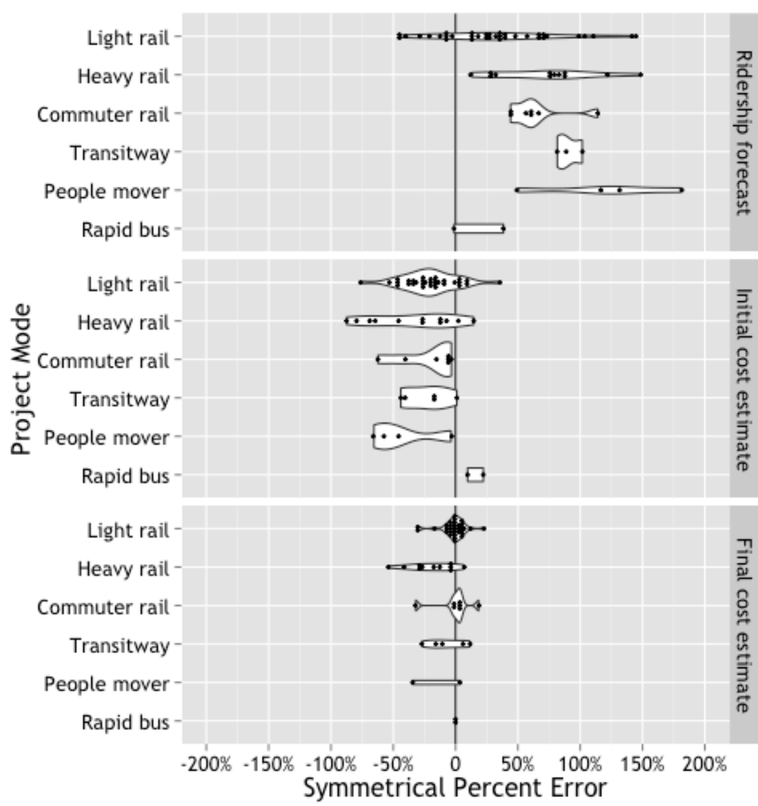

Figure 23. Variation in forecast error by project mode

Figure 23 shows the variation in forecast accuracy by project mode, and these differences are quantified in Table 18. Recall from Chapter 6 that projects of particular modes have varied in popularity over time.

The only statistically significant difference in ridership forecast accuracy by project mode is between light rail and downtown people movers. Light rail projects have final cost estimates with errors that are 18 to 154 percentage points lower than those for heavy rail projects. There are no statistically significant differences in initial cost estimate accuracy by project mode. The only statistically significant difference in final cost estimate accuracy by project mode is between light rail and heavy rail. Light rail projects have final cost estimates with errors that are four to 32 percentage points higher (i.e. less negative and more accurate) than those for heavy rail projects.

Table 18: Differences in forecast error by mode

95-percent confidence interval for difference in error (row minus column)						
	Light rail	**Heavy rail**	**Commuter rail**	**Transit-way**	**People mover**	**Rapid bus**
Ridership forecasts						
Light rail		-0.81 to 0.05	-0.84 to 0.24	-1.35 to 0.21	-1.54 to -0.18	-0.79 to 1.10
Heavy rail			-0.54 to 0.70	-1.03 to 0.65	-1.23 to 0.27	-0.46 to 1.52
Commuter rail				-1.16 to 0.63	-1.37 to 0.26	-0.59 to 1.49
Transitway					-1.28 to 0.70	-0.47 to 1.91
People mover						-0.11 to 2.13
Rapid bus						
Initial cost estimates						
Light rail		-0.11 to 0.38	-0.32 to 0.29	-0.32 to 0.38	-0.17 to 0.61	-0.90 to 0.16
Heavy rail			-0.49 to 0.20	-0.50 to 0.28	-0.34 to 0.51	-1.06 to 0.05
Commuter rail				-0.39 to 0.47	-0.22 to 0.69	-0.94 to 0.23
Transitway					-0.29 to 0.69	-1.01 to 0.22
People mover						-1.22 to 0.04
Rapid bus						
Final cost estimates						
Light rail		0.04 to 0.32	-0.17 to 0.16	-0.14 to 0.25	-0.15 to 0.43	-0.31 to 0.28
Heavy rail			-0.38 to 0.01	-0.34 to 0.09	-0.35 to 0.27	-0.50 to 0.11
Commuter rail				-0.17 to 0.30	-0.18 to 0.47	-0.33 to 0.31
Transitway					-0.25 to 0.42	-0.41 to 0.26
People mover						-0.56 to 0.25
Rapid bus						
Gray text indicates that the difference is not significant at a 95-percent confidence level.						

Variation with local experience

As described in Chapter 5, interviewees suggested that improvements in cost and ridership forecast accuracy over time might be explained by local experience with planning for fixed-guideway transit projects, such that forecasts for a project expanding or renovating an existing fixed-guideway system would use be more accurate than those for the initial line of a new system. However, this analysis does not indicate that this is the case, whether local experience is described by a categorical or a continuous variable. This finding is consistent with previous work by Schmitt (2016b), and might indicate that local conditions do not have important unexpected effects on project ridership, so that forecasters (who may be consultants who do work in a variety of locations) can easily apply lesson learned in one place to forecasts for a project elsewhere.

Project sequence

Figure 24 illustrates how forecast accuracy varies by project sequence, and these differences are quantified in Table 19. As shown, none of the differences are significant at a 95-percent confidence level.

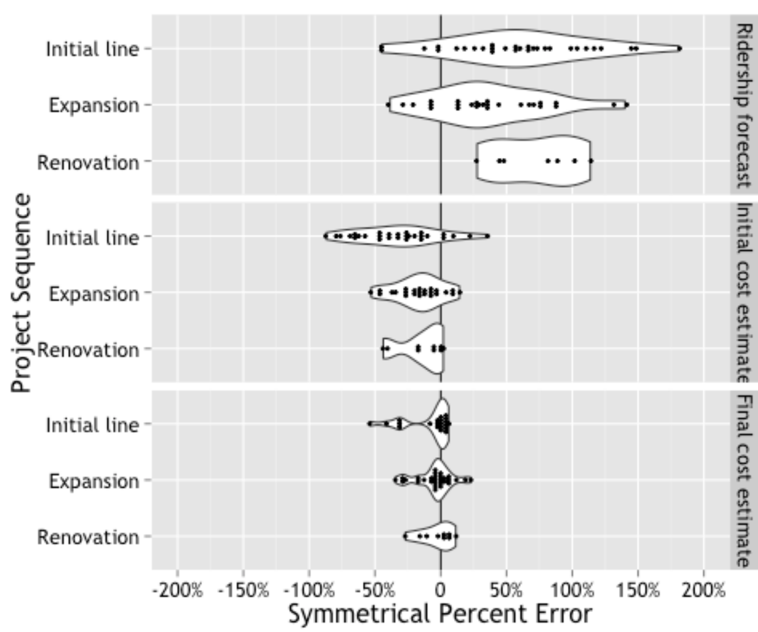

Figure 24. Variation in forecast error by project sequence

Table 19: Differences in ridership forecast error by project sequence

95-percent confidence interval for difference in forecast error (row minus column)			
	Initial line	Expansion	Renovation
Ridership forecasts			
Initial line		-0.09 to 0.53	-0.84 to 0.15
Expansion			-0.62 to 0.37
Renovation			
Initial cost estimates			
Initial line		-0.33 to <0.01	-0.41 to 0.04
Expansion			-0.26 to 0.21
Renovation			
Final cost estimates			
Initial line		-0.14 to 0.07	-0.20 to 0.09
Expansion			-0.15 to 0.12
Renovation			

Gray text indicates that the difference is not significant at a 95-percent confidence level.

System mileage share

Among the projects categorized as expansions of existing systems, there is substantial variation in the degree of the expansion: a one-mile expansion of an existing 100-mile system draws upon more relevant local experience than a 100-mile expansion of an existing 1-mile system does. A project's resulting system mileage share —calculated as the ratio of the project length to the sum of the project length and the length of the existing system at the time of the forecast— captures this variation better than a single "expansion" category can.

Using this more sensitive measure of local experience, there is a statistically significant relationship between local experience and initial cost estimate error (with a 95-percent confidence interval for the correlation between -0.50 and -0.04), but the relationships with ridership forecast error and final cost estimate error are insignificant (with respective 95-percent confidence intervals from -0.25 to 0.28 and from -0.39 and 0.12). These relationships are illustrated in Figure 25.

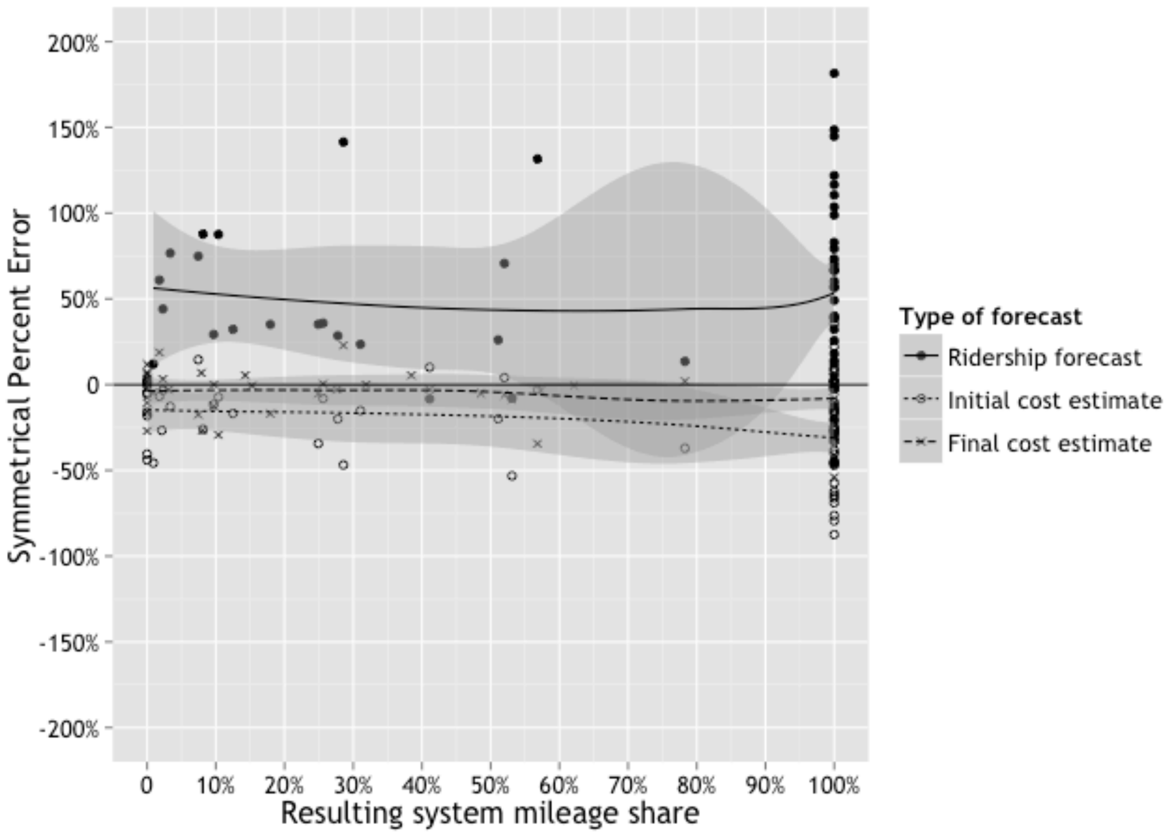

Figure 25. Variation in forecast error by resulting system mileage share

Variation with external factors

I also tested four relationships between forecast accuracy and absolute levels of population and unemployment, as well as changes in population, unemployment, and fuel price.

Population

Project sponsors serving larger populations may have greater resources to devote to preparing rigorous forecasts. They may also answer to a wider variety of stakeholders, which could influence the incentives for promoting a particular project through optimistic forecasts. However, an analysis of the relationship between county population and forecast error, as illustrated in Figure 26, does not confirm a relationship between population and forecast accuracy. Since the county populations of projects in the sample follow a lognormal distribution, the x-axis in Figure 26 uses a lognormal scale.

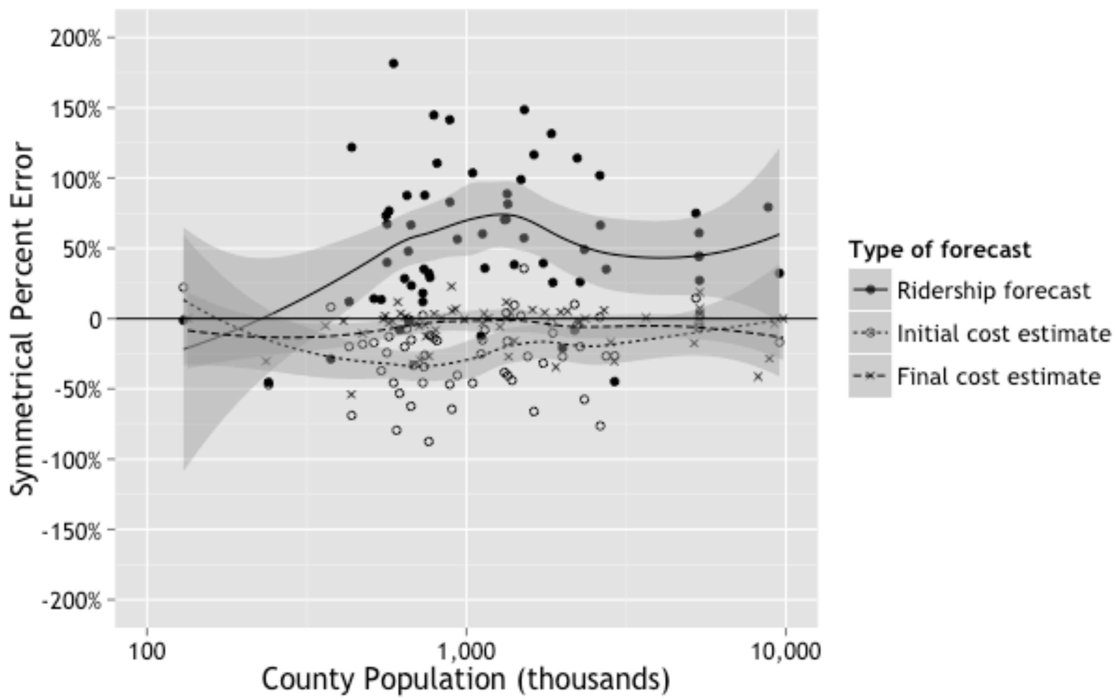

Figure 26. Variation of forecast error by county population

County population is not correlated with the errors of ridership forecasts, initial cost estimates, or final cost estimates. The 95-percent correlation for ridership forecast error is between -0.25 and 0.25; for initial cost estimates, it is between -0.01 and 0.36; and for final cost estimates, it is between -0.32 and 0.19. These correlations remain insignificant when population is log-transformed.

Unemployment rate

When unemployment is high, pressures to attract federal investment in transit projects might also be higher, which could incentivize forecast optimism. Likewise, under high unemployment, competition for construction contracts might be fierce, and a few studies have found that the competitiveness of the contract bidding environment tends to make cost overruns more likely (Rowland 1981; Jahren and Ashe 1990; Hinze, Selstead, and

Mahoney 1992). However, if high unemployment also affects the price of labor, lower labor costs could drive down construction costs down such that, even with cost overruns, actual costs are lower during periods of high unemployment than during periods of low unemployment.

As illustrated in Figure 27, there does appear to be a relationship between unemployment rates and forecast accuracy, particularly for ridership forecasts and final cost estimates. The local unemployment rate is correlated with the errors in ridership forecasts (95-percent confidence interval between 0.09 and 0.55) and final cost estimates (95-percent confidence interval between -0.56 and -0.11), but not with the errors in initial cost estimates (95-percent confidence interval between -0.45 and 0.04).

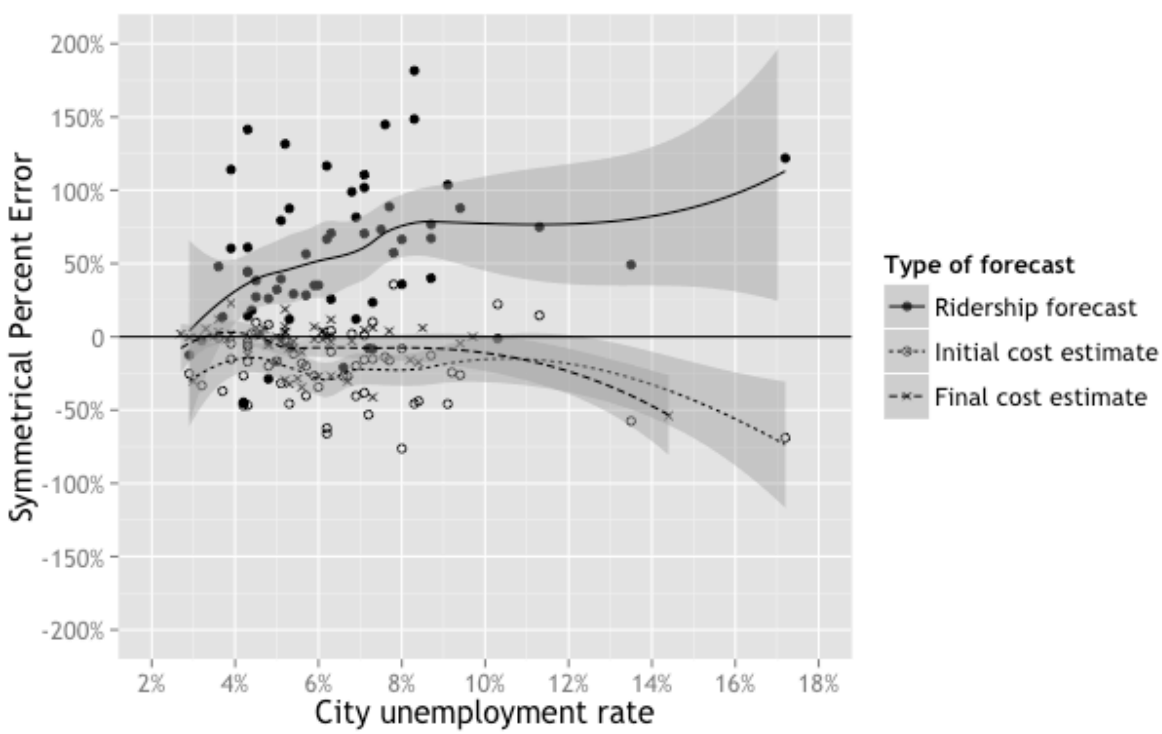

Figure 27. Variation in forecast error by city unemployment rate

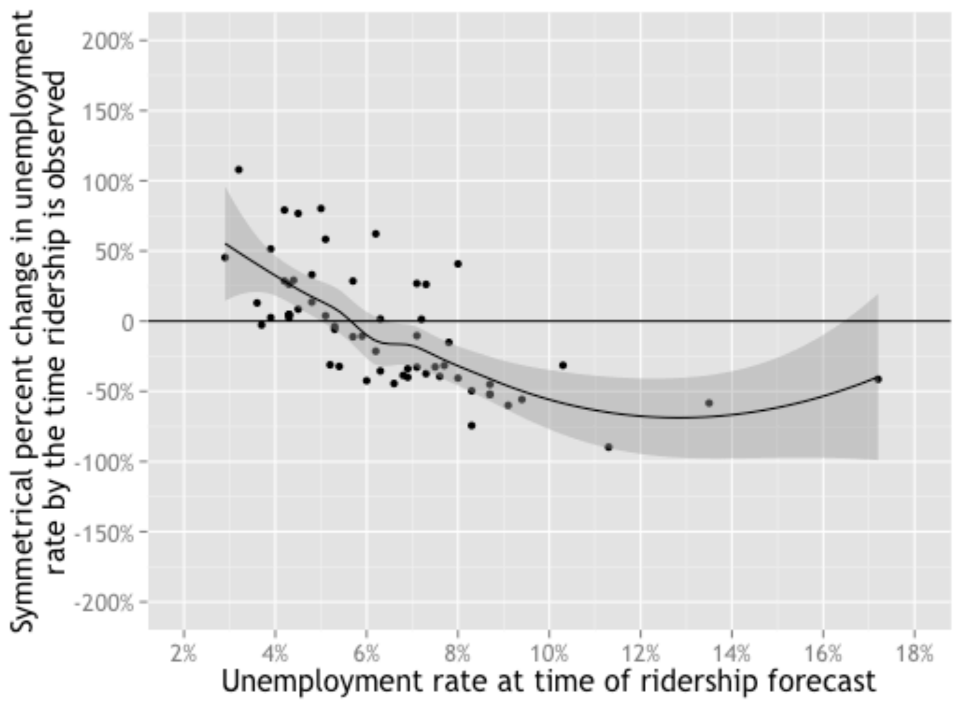

Figure 28. Relationship between level and subsequent change in unemployment

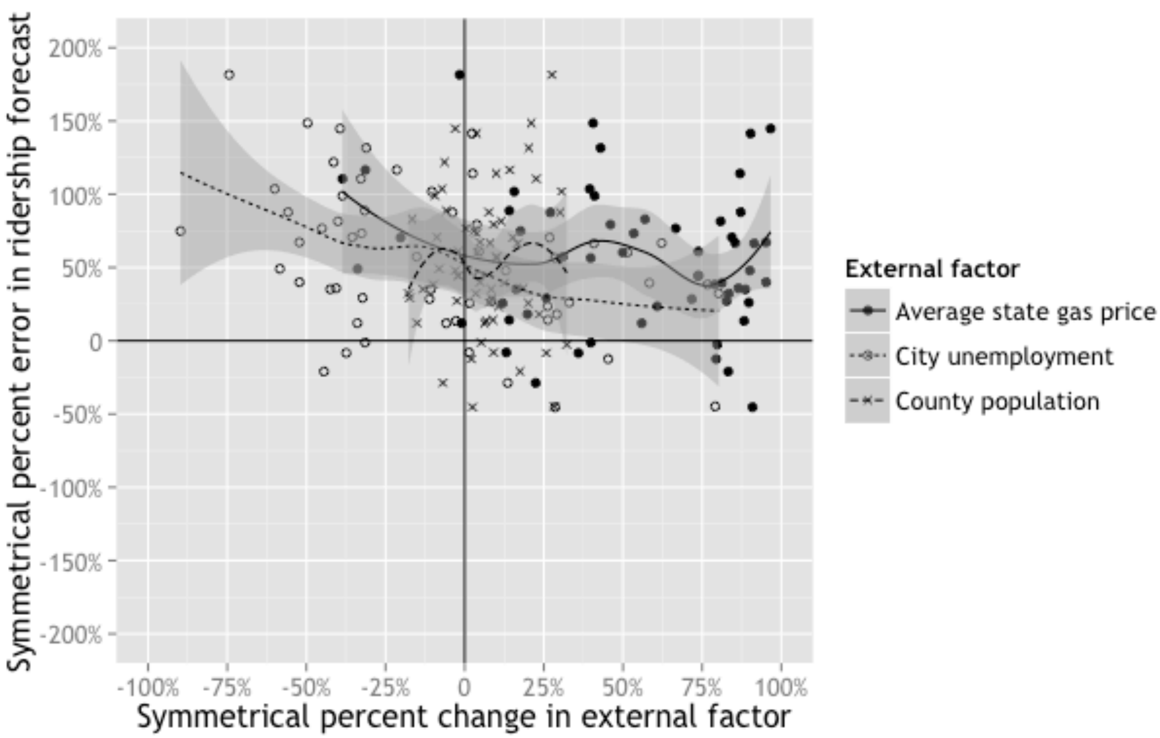

Figure 29. Variation in ridership forecast error by changes external conditions

Change in fuel price, unemployment, and population

Since ridership forecasts depend on assumptions about future population, unemployment rates, and fuel prices, dramatic changes in these values between the time the forecast is made and the time ridership is observed might be expected to be associated with less accurate ridership forecasts. Indeed, there is a significant negative correlation between the change in the unemployment rate and ridership forecast error, with a 95-percent confidence interval between -0.63 and -0.21, so that reductions in the unemployment rate are associated with less accurate ridership forecasts, and increases in the unemployment rate are associated with more accurate forecasts. As discussed above, this relationship is probably due to the positive association between unemployment rates at the time of a forecast and ridership forecast errors. As shown in Figure 28, the level of unemployment at the time of a forecast is negatively correlated with the subsequent change in unemployment (the 95-percent confidence interval for the correlation is between -0.76 and -0.45). This reflects the phenomenon of "regression to the mean" (Healy and Goldstein 1978): When the unemployment rate is high (higher than about 5.5 percent) at the time a ridership forecast is made, it will generally decrease by the time ridership is observed. Conversely, when the unemployment rate is low, it is likely to increase by the time ridership is observed.

Changes in population and fuel prices are not correlated with ridership forecast error: The 95-percent confidence intervals for those correlations are -0.20 to 0.30 and -0.43 to 0.06, respectively. Figure 29 illustrates these relationships.

Models

The analysis summarized in this chapter thus far suggests several relationships between the accuracy of three types of forecasts and project characteristics or external factors. These relationships are summarized in Table 20.

Table 20: Simple relationships with forecast accuracy

	Forecast type		
	Cost estimate		Ridership forecast
	Initial	Final	
Time of forecast			
Year	+	o	+
Governing legislation	o	o	o
Project characteristics			
Elapsed time			
Project development	o	—	o
Project construction	—	o	—
Ridership forecast to horizon year	na	na	o
Financial characteristics			
Total project cost	—	—	o
Federal funding share	+	+	o
Physical characteristics			
Project length	o	o	o
Project mode	o	□	□
Local experience			
Project sequence	o	o	o
Project share of forecast-year system miles	—	o	o
External control variables			
Change in fuel price	na	na	o
Change in unemployment	na	na	+
Change in population	na	na	o
Forecast-year unemployment	o	—	—
Forecast-year population	o	o	o

Relationship with accuracy, without controlling for other factors
\+ = *Increase associated with improved accuracy*
— = *Increase associated with reduced accuracy*
o = *No significant relationship with accuracy*
□ = *At least one pair of categories variables associated with a difference in accuracy*
na = *Relationship not tested*

A shown in Table 20, more accurate initial cost estimates are associated with later years of cost estimate preparation, shorter construction durations, lower total project costs, higher federal funding shares, and lower shares of resulting system mileage. More accurate final cost estimates are associated with shorter remaining project development times, lower total project costs, higher federal funding shares, light rail projects, and lower unemployment at the time the cost estimate is prepared. Finally, more accurate ridership forecasts are associated with later years of forecast preparation, shorter project construction durations, light rail projects, lower unemployment rates at the time of forecast preparation, and increases in unemployment by the time ridership is observed.

Since some of these factors may vary together to some degree, and may even directly influence one another, a multivariate model is useful for determining the independent relationship of each factor with forecast accuracy for each forecast type. However, multi-collinearity can be introduced when independent variables are too closely correlated. In these cases, model fit is improved when variables are selectively omitted based on an analysis of generalized variance inflation factors (GVIF) and analyst insights and hypotheses on likely processes of cause and effect. The relationship of sets of variables to forecast accuracy can be evaluated based on comparing model fit when the variables are included to when they are excluded. Once a preferred model is identified, the independent relationship of individual variables to forecast accuracy can be evaluated based on the magnitude, direction, and significance of variable coefficients.

Ridership forecast accuracy

As the left side of Table 21 shows, when all variables presented in this chapter are included in a model predicting ridership forecast accuracy, several variables have high GVIF values, where a GVIF is determined to be too high if it exceeds 2.0 when raised to the inverse power of twice the variable's degrees of freedom. Based on this result, I removed five variables (authorizing legislation, project length, project sequence, change in unemployment, and the difference between the forecast horizon year and the ridership observation year) from the model. I removed authorizing legislation and project because they are categorical variables that are perfectly predicted by continuous variables that remain in the final model. I removed project length because it is highly correlated ($r = 0.46$) with project cost, and change in unemployment because it is highly correlated ($r = -0.63$) with forecast year unemployment. As the right side of Table 21 shows, these changes result in acceptable GVIF values.

Table 21. Generalized variance inflation factor analysis for ridership error models

	Full model			Final model		
	GVIF	df	$GVIF^{(1/(2\times df))}$	GVIF	df	$GVIF^{(1/(2\times df))}$
Time of Forecast						
Year	23.7	1	4.9	3.5	1	1.9
Legislation	525.2	4	2.2	-	-	-
Project characteristics						
Elapsed time						
Project development	4.0	1	2.0	1.6	1	1.3
Project construction	4.6	1	1.9	2.0	1	1.4
Forecast horizon	30.3	1	5.5	2.1	1	1.4
Financial characteristics						
Total project cost (log-transformed)	4.7	1	2.2	3.1	1	1.7
Federal funding share	2.0	1	1.4	1.7	1	1.3
Physical characteristics						
Project length (log-transformed)	3.6	1	1.9	-	-	-
Project mode	187.2	5	1.7	11.4	5	1.3
Local experience						
System mileage share	8.5	1	2.9	2.9	1	1.7
Project sequence	19.0	2	2.1	-	-	-
External control variables						
Change in fuel price	5.9	1	2.4	2.8	1	1.7
Change in unemployment	7.1	1	2.7	-	-	-
Change in population	1.9	1	1.4	1.4	1	1.2
Forecast horizon year minus observation year	32.2	1	5.7	-	-	-
Forecast-year unemployment	4.3	1	2.0	1.6	1	1.3
Forecast-year population (log-transformed)	2.3	1	1.5	1.8	1	1.3

Figure 30 compares the fit of four different models predicting ridership forecast error, with two different measures to describe model fit: the Akaike information criterion (AIC) and adjusted R^2. Both measures are designed to avoid over-fitting a model to the data by penalizing model complexity and rewarding parsimony. The constant model predicts no variation in forecast error, and thus predicts none of the observed variation in forecast error, resulting in an adjusted R^2 value of zero. The control model includes only the time-of-forecast and external control variables listed in Table 21; the full model includes all of the variables listed in Table 21; and the final model omits variables that introduce multi-collinearity, as described above.

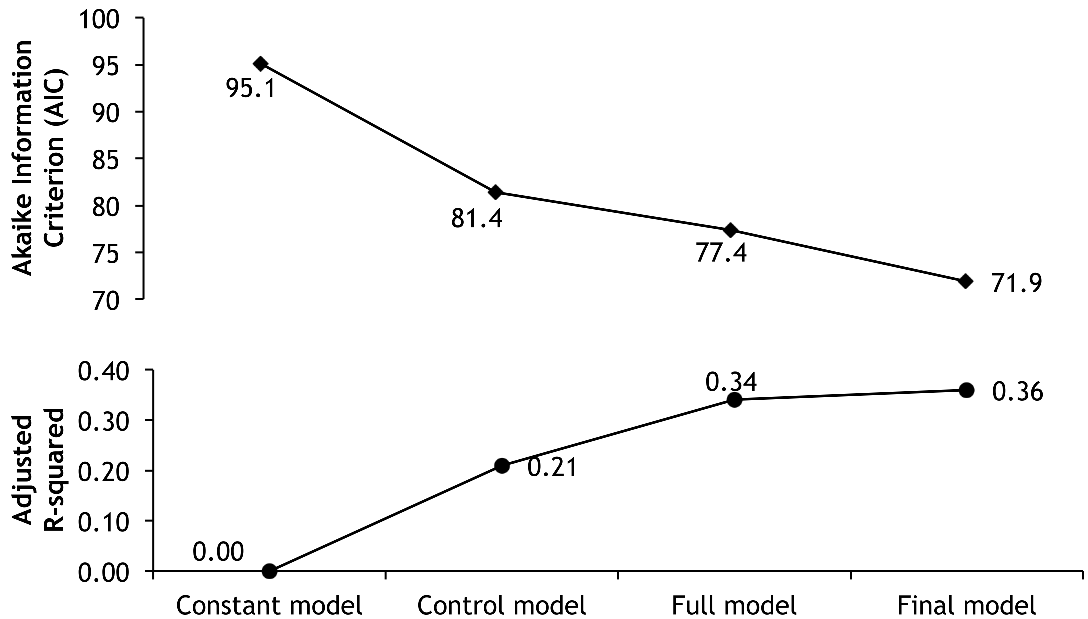

Figure 30. Model fit comparison for ridership forecast error

Figure 30 shows that the control model fits the data better than the constant model, and the full model fits the data better than the control model. This result indicates that the set of control variables and the set of variables representing project characteristics each contributes to explaining the variation in ridership forecast error. The final model offers the best model fit. The final model has an unadjusted R^2 value of 0.54, which exceeds the threshold of 0.28 required to achieve a statistical power of at least 0.80 (as discussed in Chapter 4). The statistical power of the final model is greater than 0.99, which indicates that the probability that a truly significant relationship (with a significance level of 0.05) is not reflected in the model results is less than one percent.

Table 22. Final model results for ridership forecast error

	Estimate	Standard error	p-value
Time of forecast			
Year	-0.024	0.012	0.057
Project characteristics			
Elapsed time			
Project development (years)	0.028	0.034	0.425
Project construction (years)	**0.072**	**0.029**	**0.018**
Forecast horizon (years)	-0.016	0.122	0.195
Financial characteristics			
Total project cost (log-transformed)	-0.272	0.212	0.206
Federal funding share	0.691	0.426	0.112
Physical characteristics			
Project mode (base: Light rail)			
Heavy rail	0.172	0.194	0.382
Commuter Rail	**0.425**	**0.209**	**0.049**
People mover	0.450	0.263	0.095
Bus rapid transit	-0.258	0.416	0.539
Transitway	0.272	0.294	0.361
Local experience			
System mileage share	-0.010	0.199	0.960
External control variables			
Change in fuel price	0.380	0.231	0.107
Change in population	0.800	0.512	0.128
Forecast-year unemployment	0.017	0.030	0.573
Forecast-year population (millions)	-0.009	0.182	0.961

Bold text indicates significance at a 95-percent confidence level.
Black text indicates significance at a 90-percent confidence level.
Gray text indicates non-significance.

Table 22 shows the variable coefficient estimates and p-values for the final model. As shown, the direction of the relationship with ridership forecast error can only be determined with 95-percent confidence for two variables: construction duration and the commuter rail mode. On average, and controlling for other factors, each additional year of project construction is associated with an additional seven percentage points of ridership forecast error and ridership forecasts for commuter rail projects have errors that are about 43 percentage points greater than those for light rail projects.

Two additional coefficients are significant at a relaxed significance level of 0.10: the year in which the forecast was prepared and the people mover mode. On average, and controlling for other factors, ridership forecasts for people mover projects have errors that are about 45 percentage points greater than those for light rail projects. Additionally, forecast error decreased by about 2.4 percentage points each year. This gradual improvement over time might be explained in part by the changing perspectives on the purpose of forecasting, as described in Chapter 5.

Initial cost estimate accuracy

Table 23. Generalized variance inflation factor analysis for initial cost error models

	Full model			Final model		
	GVIF	df	$GVIF^{(1/(2\times df))}$	GVIF	df	$GVIF^{(1/(2\times df))}$
Time of Forecast						
Year	21.1	1	4.6	2.4	1	1.5
Legislation	120.8	4	1.8	-	-	-
Project characteristics						
Elapsed time						
Project development	1.9	1	1.4	1.4	1	1.2
Project construction	1.7	1	1.3	1.4	1	1.2
Financial characteristics						
Total project cost (log-transformed)	3.3	1	1.8	2.6	1	1.6
Physical characteristics						
Project length (log-transformed)	2.4	1	1.5	2.1	1	1.4
Project mode	51.6	5	1.4	11.3	5	1.3
Local experience						
System mileage share	9.4	1	3.1	3.1	1	1.8
Project sequence	19.9	2	2.1	-	-	-
External control variables						
Forecast-year unemployment	1.9	1	1.4	1.4	1	1.2
Forecast-year population (log-transformed)	1.8	1	1.3	1.7	1	1.3

The full model for cost estimate accuracy includes fewer variables than the ridership model, since it does not include total forecast horizon or any of the changes in external variables. Nevertheless, as shown in the left side of Table 23, and analysis of GVIF values indicates issues with multi-collinearity, since the GVIF values raised to an inverse power of twice the degrees of freedom is greater than two for both forecast year and project sequence. Although project mode has a high GVIF value in the full model, this is not a cause for concern since the variable is a categorical variable with five degrees of freedom. The right side of Table 23 shows that the multi-collinearity is resolved in the final model when I remove the variables for authorizing legislation and project sequence. Both of these categorical variables have analogous continuous variables that remain in the final model.

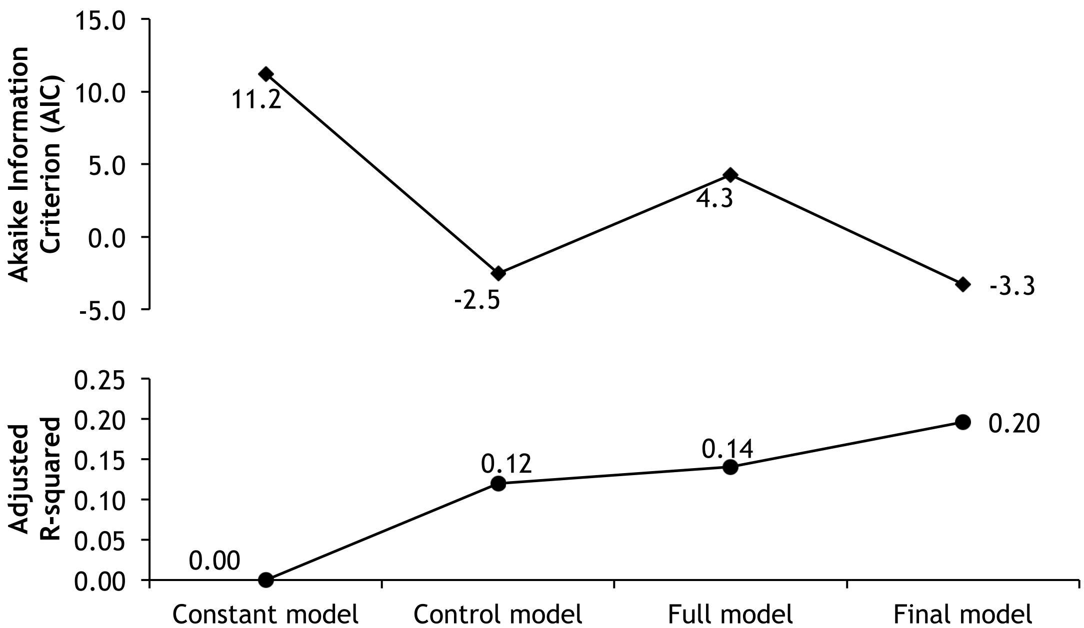

Figure 31. Model fit comparison for initial cost estimate error

Figure 31 compares model fit for predicting initial cost estimate error using the same metrics and the same four model versions that were used for predicting ridership forecast error. Based on both adjusted R^2 values and AIC, the control model fits the data better than the constant model, indicating that the control variables do contribute to explaining initial cost estimate error. However, although the adjusted R^2 value improves slightly when all of the project characteristic variables in the full model are added, the AIC score indicates that the full set of project characteristics variables degrades fit relative to the control model. The final model offers the best fit in terms of both adjusted R^2 and AIC. The final model has an unadjusted R^2 value of 0.38, which exceeds the threshold of 0.26 required to achieve a statistical power of at least 0.80. The statistical power of the final model is 0.97, which indicates that the probability of a statistically significant relationship that is not reflected in the model results is about three percent.

Table 24. Final model results for initial cost estimate error

	Estimate	Standard error	p-value
Time of forecast			
Year	-0.004	0.005	0.454
Project characteristics			
Elapsed time			
Project development (years)	0.001	0.018	0.960
Project construction (years)	-0.002	0.018	0.905
Financial characteristics			
Total project cost (log-transformed)	-0.199	0.111	0.079
Physical characteristics			
Project length (log-transformed)	0.154	0.099	0.126
Project mode (base: Light rail)			
Heavy rail	0.054	0.114	0.638
Commuter Rail	-0.150	0.112	0.188
People mover	-0.215	0.132	0.110
Bus rapid transit	0.367	0.203	0.078
Transitway	-0.091	0.130	0.489
Local experience			
System mileage share	-0.135	0.109	0.221
External control variables			
Forecast-year unemployment rate	-0.019	0.013	0.157
Forecast-year population (millions)	0.135	0.099	0.179
Bold text indicates significance at a 95-percent confidence level. Black text indicates significance at a 90-percent confidence level. Gray text indicates non-significance.			

Although the set of project characteristics included in the final model does have a collective relationship with initial cost estimate error, none of the individual variables in the final model has a statistically significant effect with initial cost estimate error at a 95-percent confidence level, as shown in Table 24. However, two coefficients are significant at a relaxed significance level of 0.10: total project cost and the bus rapid transit mode. On average, and controlling for other factors, initial cost estimates for bus rapid transit projects have errors that are about 37 percentage points more accurate than those for light rail projects. Additionally, a doubling of project cost is associated with a reduction in forecast accuracy of about 20 percentage points.

Final cost forecast accuracy

As shown in Table 25, the full model predicting final cost estimate error shows the same overall pattern of GVIF values as that predicting initial cost estimate error. Multi-collinearity issues are likewise resolved by omitting authorizing legislation and project sequence, while including their analogous continuous variables.

Table 25. Generalized variance inflation factor analysis for final cost error models

	Full model			Final model		
	GVIF	df	$GVIF^{(1/(2\times df))}$	GVIF	df	$GVIF^{(1/(2\times df))}$
Time of Forecast						
Year	29.9	1	5.5	2.6	1	1.6
Legislation	73.7	4	1.7	-	-	-
Project characteristics						
Elapsed time						
Project development	2.7	1	1.6	2.1	1	1.5
Project construction	2.4	1	1.5	1.8	1	1.3
Financial characteristics						
Total project cost (log-transformed)	3.3	1	1.8	3.3	1	1.8
Physical characteristics						
Project length (log-transformed)	2.7	1	1.7	2.2	1	1.5
Project mode	55.5	5	1.5	15.4	5	1.3
Local experience						
System mileage share	7.4	1	2.7	1.9	1	1.4
Project sequence	17.8	2	2.1	-	-	-
External control variables						
Forecast-year unemployment	2.4	1	1.6	1.7	1	1.3
Forecast-year population (log-transformed)	1.9	1	1.4	1.7	1	1.3

Figure 32 shows that model fit across the four model specifications follows the same general pattern for predicting final cost estimate error as Figure 31 shows for models predicting initial cost estimate error. The final model offers the best fit and has an unadjusted R^2 value of 0.42, which exceeds the threshold of 0.28 required to achieve a statistical power of at least 0.80. The statistical power of the final model is 0.99, which indicates that the probability of a statistically significant relationship that is not reflected in the model results is about one percent.

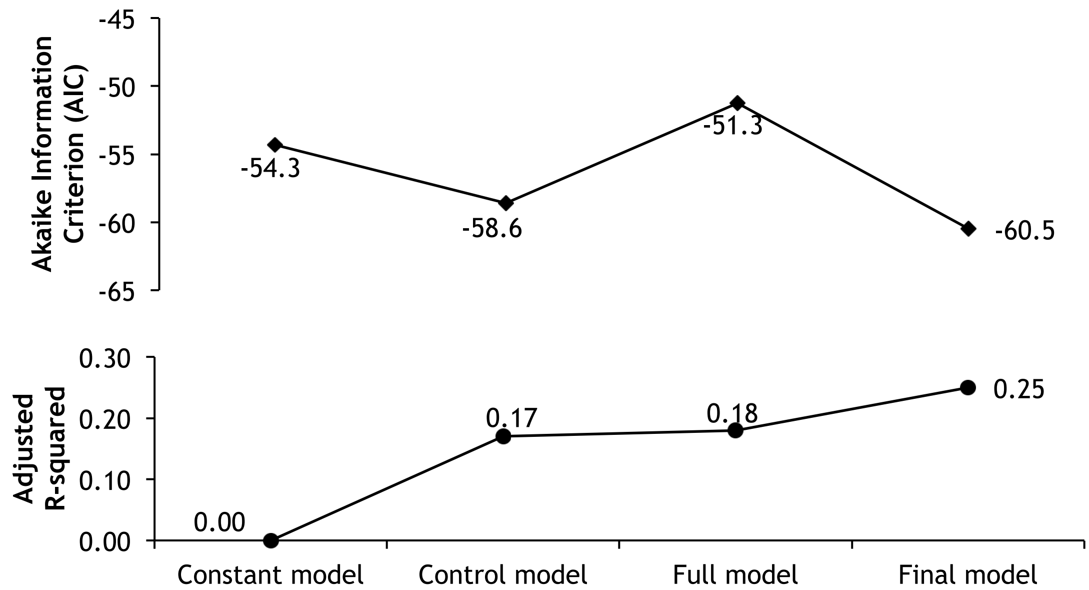

Figure 32. Model fit comparison for final cost estimate error

Table 26. Final model results for final cost estimate error

	Estimate	Standard error	p-value
Time of forecast			
Year	> -0.001	0.004	0.920
Project characteristics			
Elapsed time			
Project development (years)	-0.010	0.032	0.747
Project construction (years)	0.005	0.010	0.629
Financial characteristics			
Total project cost (log-transformed)	-0.127	0.071	0.078
Physical characteristics			
Project length (log-transformed)	0.031	0.063	0.625
Project mode (base: Light rail)			
Heavy rail	**-0.137**	**0.067**	**0.045**
Commuter Rail	-0.070	0.069	0.319
People mover	-0.152	0.124	0.227
Bus rapid transit	-0.005	0.131	0.971
Transitway	-0.079	0.083	0.378
Local experience			
System mileage share	-0.052	0.054	0.342
External control variables			
Forecast-year unemployment rate	-0.016	0.012	0.166
Forecast-year population (millions)	0.066	0.054	0.226

Bold text indicates significance at a 95-percent confidence level.
Black text indicates significance at a 90-percent confidence level.
Gray text indicates non-significance.

Table 26 shows the variable coefficient estimates and p-values for the final model. As shown, the direction of the relationship with ridership forecast error can only be determined with 95-percent confidence for the heavy rail mode. On average, final cost estimates for heavy rail projects have errors that are about 14 percentage points less accurate than those for light rail projects. At a relaxed significance level of 0.10, total project cost is also significant. On average, and controlling for other factors, a doubling of project cost is associated with a reduction in accuracy of about 13 percentage points.

Conclusion

The analyses presented in this chapter show that there have been steady improvements over time in the accuracy of ridership forecasts and initial cost estimates for rapid transit construction projects in the U.S., as indicated by the negative correlations between error magnitude and the year in which a forecast was prepared. For initial cost estimates, this relationship is not statistically significant in a model controlling for project characteristics and external factors. This finding suggests that the improvements in cost estimate accuracy over time can largely be explained by changes in project characteristics, particularly a tendency toward lower-cost projects. For ridership forecasts, the relationship between forecast error and the year in which a forecast was prepared is significant at a 90-percent confidence level, even when controlling for other project characteristics. Some of this gradual improvement over time might be attributable to changes in perspectives about the purpose of New Starts forecasts, as I describe in Chapter 5.

Based on the regression analysis, the most important project characteristic related to finding ridership forecast accuracy is the length of time required to construct the project. This suggests, quite intuitively, that the further into the future a project comes online, the greater the uncertainty at the time of the forecast. The Jacksonville Skyway Express downtown people mover offers the most extreme example of this relationship. This project was constructed over a period of 13 years (compared to a median construction duration of four years for all other projects in the study sample, and a planned construction duration of approximately three years), and only achieved five percent of forecast ridership (representing a symmetrical percent error of 182 percent, compared to a median error of 52 percent). Some of the large error might also be explained by characteristics of the downtown people mover mode, which is likewise associated with less accurate forecasts at a 90-percent confidence level.

None of the variables included in the model predicting the accuracy of initial cost estimates are significant at a 95-percent confidence interval, although two variables are associated with initial cost estimate accuracy at a 90-percent confidence level: project cost and the bus rapid transit mode. Taking these results together with the direction of individual correlations between each of these variables and initial cost estimate accuracy, it appears that the most accurate cost estimates are for projects that represent incremental changes to the existing transit network, as indicated by their low costs, short construction durations, and relatively small increases in total system mileage. Nine of the ten projects in the sample with the least accurate initial cost estimates were initial lines for new fixed-guideway systems. The tenth was the Hillsboro light rail extension in Portland, which approximately doubled the mileage of the existing system.

Final cost estimates have generally been the most accurate and least biased of the three forecast types evaluated in this study, with a median optimism error of less than one percent (compared to 20 percent for initial cost estimates and 52 percent for ridership forecasts). The variable most closely associated with final cost estimate accuracy is project mode, with the most accurate cost estimates being for light rail projects. Indeed, the median error for final cost estimates for light rail projects is zero (compared to a three-percent optimism bias for projects of other modes). More accurate final cost estimates are also associated (at a 90-percent confidence level) with lower-cost projects.

The problem of optimism bias in patronage and cost forecasts for new rail transit projects has been a serious and ongoing problem. However, this analysis of forecast bias in the U.S. up to the present gives us reasons for optimism regarding the future of optimism bias in cost and ridership forecast accuracy, since forecasts appear to be on a long-term trajectory toward more accuracy and less bias. Chapter 8 discusses these results further and relates them to the findings presented in Chapters 5 and 6.

Chapter 8. Conclusion

To return to the questions motivating this dissertation, how have changes in federal policies and transit planning practice influenced the accuracy of transit forecasts? Do current methods of forecasting ridership and costs for transit infrastructure projects resemble mythical crystal balls or the metaphorical black boxes that the general public may perceive them to be? Or have improvements in accuracy and transparency lessened the aptness of these metaphors?

As discussed in Chapter 3, forecast accuracy is only one of many characteristics of a good forecast. For example, a good forecast must also be useful and be based on sound methodology and judgment. However, while accuracy is not the only criterion by which we can evaluate forecasts, it is an important one.

The large errors in predicting both cost and ridership in the early history of the New Starts program, and the continuing bias in such forecasts, lay to rest any misapprehension that the transit forecasting profession has access to some sort of genuine crystal ball that allow forecasters to perfectly predict future conditions. Nevertheless, as described in Chapter 5, some transit professionals have expressed frustration that forecasts may be held to a standard that could only be met though supernatural abilities. As one interviewee expressed:

> Nobody did the math wrong. Nobody missed the quantity. The price of oil or the price of concrete went up due to external factors that they had no control over. Somebody with a crystal ball could say there's always a risk of that. … Sometimes I want to say, 'Well, do you expect me to know what color the cars are going to be, making the turn on the road?' It's a mathematical tool. It's not perfect. — *Interview 11*

However, the lack of a crystal ball does not necessitate a black box, and some, if not all, transit professionals have welcomed the opportunity to illuminate models that had come to be seen as black boxes by members of the public. Perhaps the best example of this was the introduction of the Summit tool, which was introduced by the FTA to allow modelers to calculate travel time savings, and also dramatically improved the transparency of ridership forecasting models.

> The Summit software … was the equivalent of shining a light into a really dark box. And all of a sudden, the ridership stuff got a lot more rigorous. That produced its own reaction as well: Folks weren't happy that they had to fix all this stuff. … I don't pretend my perspective is the general perspective, but from my perspective, … we were all a whole lot better off in terms of seeing where there were mistakes happening. — *Interview 8*

For cost estimation, this sort of increased transparency has been achieved through the development of the project management oversight program, and transit professionals appear to appreciate how federal oversight has served to likewise "shine a light into a dark box" — increasing transparency and improving planning practice.

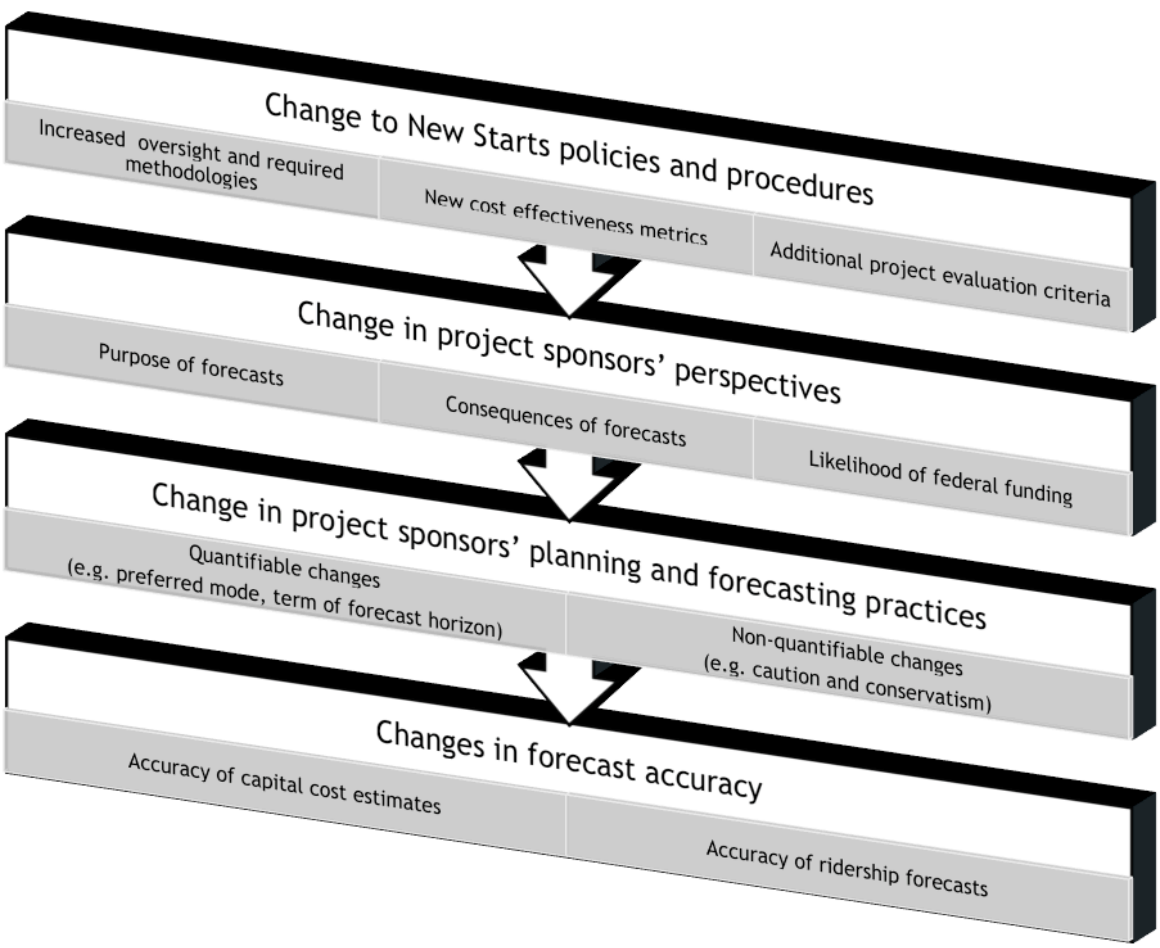

Figure 33: Hypothesized relationship between New Starts policies and forecast accuracy

The observation that improved transparency has improved planning practice points back to the hypothesis presented in Chapter 4, and illustrated in Figure 2: That changes to New Starts policies and procedures over the years have led to changes in project sponsors' perspectives on forecasting; that these changes in perspective have in turn led to changes in planning and forecasting practice; and that these changes in practice have led to changes in forecast accuracy.

How well does the analysis presented in this dissertation support these propositions? Chapter 2 describes the many changes to New Starts policies and procedures that have taken place over time, and Chapter 5 summarizes the results of interviews in which transit professionals reflected on changes they have observed in the New Starts program over the years and their perceptions of the consequences of those changes. Among the specific policies that appear to have been most influential are the use of the Summit software package between 2007 and 2012 and development of the project management oversight program, as discussed above. One other influential change that has likewise reduced the perception of forecasting models of black boxes through increased transparency has been the requirement for *ex post* analyses of cost and ridership forecasts through Before and After Studies.

Based in the results of the interviews described in Chapter 5, some of these changes in policy appear to have led to changes over time in perspectives about the purpose of preparing cost and ridership forecasts. For example, the interview results collectively suggest a shift in the perceived purpose of cost and ridership forecasts for New Starts project over time: from overcoming skepticism about the value of federal investment, to ensuring fair competition among competing project proposals, to producing information that can be useful in local decision making. Other consequences of changes in the New Starts program the interviewees discussed relate to changes in the practice of transit planning and forecasting. For example, interviewees described how particular methods of calculating cost effectiveness favored particular modes at various points in time; they explained how policies and procedures encouraged or required the use of longer or shorter forecast horizons; they noted that projects today are more likely to be expansion of existing systems than initial lines of new systems; and they mentioned that federal funding is more scarce today than in the past.

Some of the changes in practice alluded to in Chapter 5 can be quantified and observed, and Chapter 6 presents a quantitative analysis of some these changes, and confirmation of several of them. There are indeed distinct periods of time in which projects comprised of a particular mode became more or less common, and there was a steady increase in the length of ridership forecast horizons until the 1990s (the ISTEA years), followed by a sharp decline thereafter. On the other hand, the quantitative analysis failed to confirm other observations made by interviewees. For example, initial lines of new projects have continued to represent a large share of New Starts projects, and perhaps most surprisingly, federal funding shares have been relatively consistent over time. Failure to observe these trends quantitatively does not necessarily mean that interviewees were mistaken. It is possible that the effects they describe have happened too recently to be included in the dataset, that the magnitude of the effects has been too small to achieve statistical significance, or that they are hidden by selection bias within the sample. For instance, greater scarcity in federal funds may not be reflected in a smaller share of proposed projects receiving funding, rather than lower federal funding shares for those that do receive grants. However, the study sample only includes those projects that were ultimately constructed. In addition to the changes in practice that were mentioned by interviewees in Chapter 5, another significant change in practice has been a trend toward shorter project construction durations.

Have the changes in perspective and practice described above influenced forecast accuracy, as the hypothesis illustrated in Figure 2 would suggest? Chapter 7 presents an analysis of the relationships between quantifiable changes in transit planning practice and forecast accuracy. Of the three project characteristics that Chapter 6 identifies as having a significant trend over time (construction duration, mode, and ridership forecast horizon), two in particular —construction duration and mode— have a statistically significant relationship with forecast accuracy without controlling for other factors. Light rail projects are associated with more accurate final cost estimates; shorter construction durations are associated with more accurate initial cost estimates; and both are associated with more accurate ridership forecasts. When the models I estimated control for other factors, the relationship between construction duration and accuracy persists for ridership forecasts. Project mode also has a relationship with forecast accuracy for all three forecast types when it is included in models with controls for other project characteristics. Additionally, shorter construction durations and later forecast preparation years are associated with more accurate ridership forecasts, and lower-cost projects are associated with more accurate initial and final cost estimates.

I have also hypothesized that the changes in perspective about the purpose of forecasts described in Chapter 5 should correspond to a gradual improvement in forecast accuracy over time, and the analysis in Chapter 7 supports this, demonstrating a significant, positive correlation between the year in which a forecast was prepared and the accuracy of initial cost estimates and ridership forecasts. For ridership forecasts, this relationship persists even when controlling for other factors.

Do these results indicate that forecasts are *better* than they have been in the past? The answer to this question depends on the criteria chosen to evaluate forecasts. Recall from Chapter 3 that forecasting theorists (Ascher 1979; Tetlock 2009; Murphy 1993; Kahneman 2013) have suggested three distinct criteria for evaluating forecasts: methodology and judgment, accuracy, and usefulness. These are summarized in Table 2. How do the changes I have described in transit forecasting relate to each of these three criteria?

The quantitative analysis in this dissertation takes Ascher's (1979) "outsider's approach" by evaluating forecast accuracy rather than considering the specific methodologies used to produce each of the forecasts in the study sample. Thus, the quantitative analysis does not offer any insight into changes in methodology and judgment over time. However, the qualitative interview results do suggest that there has been improvement in this respect.

Murphy (1993) describes his test for forecaster judgment in terms of *consistency* between a published forecast and the forecaster's own beliefs about future conditions. Indeed, a few interviewees described how recent shifts in perspective about the role of forecasts in improving decision-making and infrastructure investment at the local level have emboldened them to resist pressure (and in some cases, experience reduced the pressure) to present forecasts that are more optimistic than their own beliefs would justify. This represents an improvement in terms of Murphy's (1993) *consistency* criterion.

Table 27: Categories of criteria for evaluating forecasts

Publication	Types of forecast addressed	Proposed criteria for forecast evaluation		
		Methodology and judgment	***Accuracy***	***Usefulness***
Ascher, William. 1979. *Forecasting: An Appraisal for Policy-Makers and Planners*. Baltimore: Johns Hopkins University Press.	Population, economic, energy, transportation, and technology	Insider's approach	Outsider's approach	*Discussed, but not an evaluation criterion*
Tetlock, Philip E. 2009. *Expert Political Judgment: How Good Is It? How Can We Know?* Kindle Edition. Princeton University Press.	Political events	Logical coherence and process tests	Empirical correspondence tests	*Discussed, but not an evaluation criterion*
Murphy, Allan H. 1993. "What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting." *Weather and Forecasting* 8 (2): 281–93.	Weather	Consistency	Quality	Value
Kahneman, Daniel. 2013. *Thinking, Fast and Slow*. Reprint edition. New York: Farrar, Straus and Giroux.	Project Planning	Inside view	Outside view	*Not discussed*

Tetlock (2009) takes Murphy's (1993) *consistency* criterion further by suggesting that a published forecast should not only reflect a forecaster's beliefs, but that those beliefs should be justified and updated based on new evidence. For transit forecasts, such evidence is more available today than it has been in the past through the publication of Before and After Studies for completed New Starts projects. Multiple interviewees mentioned that they had used the results of published Before and After Studies to better understand the overall level of uncertainty associated with their own forecasts, and also to improve their own forecasting methods (both of which are a primary purposes of Before and After Studies), although they expressed doubts that such use of Before and After Studies had been widely adopted within profession.

Given these findings, how might cost and patronage forecasting be improved further in the future? The FTA could work to improve forecaster judgment by incorporating the results of completed Before and After Studies into guidance documents and training programs. However, this information is already widely available (if not included in guidance documents) through published Before and After Studies, annual summaries of Before and After Studies, and occasional reports that aggregate case studies of forecast accuracy (Spielberg et al. 2003; Lewis-Workman et al. 2008). Thus, at least some of the onus for using the information contained in these existing reports must rest with forecasting professionals.

With regard to forecast methodology, some observers have suggested that improvements in forecast accuracy in recent years can be explained by increased computing power that facilitates greater model complexity (Spielberg et al. 2003). The analysis presented in this dissertation does not directly address the potential relationship between model complexity and forecast accuracy. However, Ascher (1979) suggests that, for forecasting in general, model complexity is not generally associated with forecast accuracy and Alonso (1968) demonstrates mathematically that model complexity will tend to exacerbate any uncertainty associated with model inputs. A few interviewees similarly expressed the view that simple models focused on addressing specific questions might ultimately offer better results than complex models that can address a wide variety of questions. The recent introduction of the STOPS software package, which was developed by the FTA as an alternative to regional travel demand models for project-level ridership forecasts, may thus represent an improvement in transit ridership forecasting methodology. If this is the case, the FTA should continue to encourage widespread adoption of STOPS or other simple, standardized modeling methods that are specifically focused on predicting transit ridership.

Moving from methodology and judgment to forecast accuracy, the quantitative analysis presented in Chapter 7 makes a clear case that forecasts are improving according to this criterion — particularly in the cases of ridership forecasts and initial cost estimates. However, there continues to be persistent optimistic bias in both initial cost estimates and ridership forecasts.

Flyvbjerg (2006, 2008) has suggested reference class forecasting as a means of overcoming bias and incorporating Kahneman's (2013) "outside view," which entails incorporating information about similar completed projects into forecasts about the performance of proposed projects. Schmitt (2016b) has proposed a method to introduce the practice of reference class forecasting to ridership forecasts by adjusting initial forecasts by factors derived from average errors that have been observed for similar projects. The quantitative results presented in this study could be used to generate factors for use in these types of *post-hoc* adjustments to cost estimates and ridership forecasts.

For example, based on my findings that final cost estimates for heavy rail projects have errors that are, on average 13.7 percentage points less accurate than those for light rail projects, and that final cost estimates for light rail projects are unbiased, a forecast for a proposed heavy rail project might be adjusted based on the equation for symmetrical mean percent error (Equation 3), which can be rearranged and rewritten as Equation 4, where *i* represents the unadjusted final cost estimate and *a* represents the adjusted final cost estimate. However, care should be taken in applying this approach, since it is based on a sample of many projects that may be very different from the case to which it is applied. Comparison to a smaller number of very similar projects is likely to be more helpful than applying factors based on a the full sample included in this study.

Equation 3:

$$\frac{(i-a)}{\left(\frac{1}{2}\right)(i+a)} = 0.137$$

Equation 4:

$$a = i\left(\frac{1+\left(\frac{1}{2}\right)0.137}{1-\left(\frac{1}{2}\right)0.137}\right)$$

At any rate, simply adjusting a point forecast is not sufficient to ensure forecast accuracy. Murphy (1993) points out that a high quality forecast should faithfully represent both an expected value and the uncertainty associated with that value. Thus, while the type of adjustment described above might serve to reduce the magnitude of over- and under-estimates, they would not produce truly accurate forecasts unless the forecasts also incorporated information about the uncertainty surrounding forecast values. Pickrell (1992) has suggested that forecasts for transit projects be presented as ranges of likely values rather than as point estimates. The results of this study indicate that forecast errors continue to vary widely, which supports the idea that ranges of forecast values would be valuable, particularly if they were associated with a probability that the actual value would fall within the forecast range. Moreover, multiple interviewees acknowledged the uncertainty associated with forecasting and expressed a preference for presenting forecasts in terms of ranges rather than as single points. However, although it has been a quarter century since Pickrell (1992) put this suggestion forward, and the idea appears to have the support of more than a few transit professionals, it has yet to be widely adopted as a standard practice in cost and ridership forecasting for transit projects.

The reason the practice of presenting cost and ridership forecasts in terms of ranges has not been widely adopted relates in part to the third criterion for forecast evaluation suggested by forecasting theorists (Ascher 1979; Tetlock 2009; Murphy 1993; Kahneman 2013). In addition to being prepared with sound judgment and methods and being accurate, a forecast must also be *useful* to forecast users.

The initial and primary users of New Starts forecasts are the FTA staff who evaluate projects for funding. Current evaluation methods require projected performance metrics (such as cost effectiveness) to exceed defined thresholds. When a forecast range includes the threshold, it creates ambiguity. In this case, a point forecast may be more useful, if ultimately less accurate, than a range of possible values.

One way to address this issue would be for evaluators to define an acceptable level of risk that a project could fail meet a defined performance threshold. A risk-based standard would not be unprecedented in transportation planning and engineering practice. Traffic engineers commonly design for the 85th percentile, accepting a 15-percent risk that design speeds (Lave 1985) and parking capacities (McCourt 2004) will be exceeded. Assigning project ratings based on the thresholds that proposed projects would have an 85-percent probability of meeting would allow for greater transparency in presenting cost and ridership forecasts.

Overall, the analyses presented in this dissertation support my initial hypothesis that changes in the New Starts program over the years, such as increased oversight and a greater emphasis on *ex post* analysis, have led to changes in perspectives on transit planning and forecasting, to changes in planning and forecasting practice, and ultimately to improvements in forecast accuracy. In order for this trend of improvement to continue, I recommend that administrators of the New Starts program and other federal, state, and local programs that fund transit infrastructure projects consider

- Widely disseminating the results of *ex post* analysis of forecasts for completed projects,
- Encouraging the development and use of models that target the specific questions that are of primary interest to project evaluators,
- Adjusting forecasts for projects with characteristics that are known to be associated with increased forecast error, and
- Adopting project evaluation procedures that can accommodate ranges of forecast values and associated levels of uncertainty.

These actions would have the potential to continue to improve forecasting practice in terms of judgment and methodology, accuracy, and usefulness. While we may never achieve the ideal of a forecasting crystal ball, we do not need to settle for a black box.

<u>**Appendix A: Interview Questions**</u>

The specific questions varied depending on the background, experience, and expertise of the interviewee. The questions below serve to illustrate the nature of the questions.

1. Tell me about your professional background in the public transit industry.
2. What has been your experience with the New Starts program?
3. Can you give an overview of the New Starts process and what it looks like on your end?
4. Can you tell me about the process of preparing an application for entry into preliminary engineering?
    a. What is your role in that process?
    b. How long before entry into preliminary engineering are cost and ridership forecasts typically prepared?
5. Can you tell me about the process of preparing an application for entry into final design?
    a. What is your role in that process?
    b. How long before entry into final design are cost and ridership forecasts typically prepared?
6. Can you tell me about the process of preparing an application for entry into a full-funding grant agreement?
    a. What is your role in that process?
    b. How long before entry into a full-funding grant agreement are cost and ridership forecasts typically prepared?
7. Can you tell me about the process of preparing a Before and After Study?
    a. What is your role in that process?
8. Is there a project you have worked on where you were unhappy with the outcome?
    a. What do you feel went wrong with that project?

9. Is there a project you have worked on that you are particularly proud of?
    a. What contributed to the success of that project?
10. With respect to your particular job duties, how do you define success?
    a. What makes a forecast a success?
    b. What makes a transit project a success?
11. What changes to the New Starts program over the years have been the most important?
    a. Why?
    b. How have these changes affected your own approach to cost and ridership forecasting?
    c. How have these changes affected the approach of others within your organization to cost and ridership forecasting?
    d. How have these change affected your overall organization's approach to cost and ridership forecasting?
    e. How do you think these changes have affected the approach of other organizations to forecasting?
12. Can you think of a time when *your organization's* approach to forecasting has changed in response to a recent policy change?
    a. How did forecasting practice within your organization change?
    b. How long after the policy changed did they begin to do things differently?
13. Can you think of a time when *your own* approach to forecasting has changed in response to a recent policy change?
    a. How did your own forecasting practice change?
    b. How long after the policy changed did you begin to do things differently?

14. Can you think of a time when *your organization's* approach to forecasting has changed in anticipation of future policy changes?
    a. How did forecasting practice within your organization change?
    b. How long before the policy change did they begin to do things differently?
15. Can you think of a time when *your own* approach to forecasting has changed in anticipation of future policy changes?
    a. How did your own forecasting practice change?
    b. How long before the policy change did you begin to do things differently?
16. Is there anyone else who might have a relevant perspective on cost and ridership forecasting for New Starts projects that you would recommend I speak to?

Appendix B: Data and data sources used for quantitative analysis

Table B-1: Time variables for study sample (1/3)

	Year of …							Funding legislation time period		
Project name	**Initial cost forecast**	**Ridership forecast**	**Final cost forecast**	**Construction start**	**Project opening**	**Ridership forecast horizon**	**Observed ridership**	**Initial cost forecast**	**Ridership forecast**	**Final cost forecast**
Heavy Rail (Baltimore, MD)	1972	1974	NA	1976	1983	1980	1987	UMTA	UMTA	NA
Heavy Rail (Washington, DC)	1969	1969	NA	1969	1985	1977	1986	UMTA	UMTA	NA
Heavy Rail (Miami, FL)	1978	1977	NA	1980	1985	1985	1988	UMTA	UMTA	NA
Heavy Rail (Atlanta, GA)	1971	NA	NA	1975	1986	NA	1987	UMTA	NA	NA
Light Rail (Buffalo, NY)	1977	1977	NA	1979	1986	1995	1989	UMTA	UMTA	NA
Metromover (Miami, FL)	1980	1980	NA	1983	1986	1985	1988	UMTA	UMTA	NA
Banfield Transitway (Portland, OR)	1980	1977	NA	1982	1986	1990	1989	UMTA	UMTA	NA
Light Rail (Pittsburgh, PA)	1978	1978	NA	1980	1987	1985	1989	UMTA	UMTA	NA
People Mover (Detroit, MI)	1980	1980	NA	1983	1987	1985	1988	UMTA	UMTA	NA
Light Rail (Sacramento, CA)	1981	1981	NA	1984	1987	2000	1989	UMTA	UMTA	NA
El Cajon Extension (San Diego, CA)	1985	1985	1986	1987	1989	2000	2000	UMTA	UMTA	UMTA
Downtown Transit (Seattle, WA)	1985	NA	1986	1986	1990	NA	1991	UMTA	NA	UMTA
Guadalupe (San Jose, CA)	1981	1981	1984	1986	1991	1990	1992	UMTA	UMTA	UMTA
Orange Line (Chicago, IL)	1982	1982	1986	1987	1993	2000	2000	UMTA	UMTA	UMTA
SW Transitway (Houston, TX)	1985	1985	1987	1987	1993	2005	2002	UMTA	UMTA	UMTA
Metrolink (St Louis, MO)	1984	1984	1988	1990	1994	1995	1995	UMTA	UMTA	STURAA
Metromover Extensions (Miami, FL)	1987	1987	1989	1991	1994	2000	2000	UMTA	UMTA	STURAA
I-25 Transitway (Denver, CO)	1989	NA	1989	1991	1994	NA	2000	STURAA	NA	STURAA
Metro Extension (Baltimore, MD)	1984	1984	1988	1988	1995	2005	2001	UMTA	UMTA	STURAA
Colma station (San Francisco, CA)	1988	1988	1993	1993	1996	2000	2000	STURAA	STURAA	STURAA
South Oak Cliff (Dallas, TX)	1990	1990	1993	1993	1997	2005	2002	STURAA	STURAA	STURAA
LRT Extensions (Baltimore, MD)	1991	1991	1994	1994	1997	2005	2001	STURAA	STURAA	STURAA
Hillsboro Extension (Portland, OR)	1993	1993	1994	1994	1998	2005	2002	STURAA	STURAA	STURAA
N/S LRT (Salt Lake City, UT)	1990	1990	1995	1995	1999	2010	2002	STURAA	STURAA	STURAA
Tasman West LRT (San Jose, CA)	1991	1991	1996	1996	1999	2005	2002	STURAA	STURAA	STURAA
Skyway Express (Jacksonville, FL)	1982	1982	1985	1987	2000	1995	2000	UMTA	UMTA	UMTA
Red Line (Los Angeles, CA)	NA	1989	NA	1986	2000	2000	2001	NA	STURAA	NA

Table B-1: Time variables for study sample (2/3)

	Year of …							Funding legislation time period		
Project name	Initial cost forecast	Rider- ship forecast	Final cost forecast	Construction start	Project opening	Ridership forecast horizon	Observed ridership	Initial cost forecast	Ridership forecast	Final cost forecast
Red Line MOS 1 (Los Angeles, CA)	NA	NA	1985	1986	1993	NA	NA	NA	NA	UMTA
Red Line MOS 2 (Los Angeles, CA)	NA	NA	1990	1990	1999	NA	NA	NA	NA	STURAA
Red Line MOS 3 (Los Angeles, CA)	NA	NA	1997	1994	2000	NA	NA	NA	NA	ISTEA
North Line Extension (Atlanta, GA)	1990	1990	1994	1995	2000	2005	2003	STURAA	STURAA	ISTEA
Airport Busway (Pittsburgh, PA)	1992	1992	1994	1994	2000	2005	2002	ISTEA	ISTEA	ISTEA
Southwest Light Rail (Denver, CO)	1994	1994	1999	1999	2000	2015	2002	ISTEA	ISTEA	TEA-21
University Ext (Salt Lake City, UT)	NA	1999	2000	2000	2001	2020	2005	NA	TEA-21	TEA-21
St Clair Extension (St Louis, MO)	1994	1994	1996	1996	2001	2010	2002	ISTEA	ISTEA	ISTEA
North Central Ext (Dallas, TX)	1994	1994	1999	1999	2002	2020	2007	ISTEA	ISTEA	TEA-21
BART to SFO (San Francisco, CA)	1993	1993	1997	1999	2003	2010	2007	ISTEA	ISTEA	ISTEA
Med Center Ext (Salt Lake City, UT)	NA	1999	2002	2002	2003	2020	2005	NA	TEA-21	TEA-21
South LRT (Sacramento, CA)	1995	1995	1997	1997	2003	2015	2007	ISTEA	ISTEA	ISTEA
Piers Transitway (Boston, MA)	1993	1993	1994	1994	2004	2010	2007	ISTEA	ISTEA	ISTEA
Largo Extension (Washington, DC)	1996	1996	2000	2001	2004	2020	2007	ISTEA	ISTEA	TEA-21
Interstate MAX (Portland, OR)	1996	1996	2000	2000	2004	2015	2007	ISTEA	ISTEA	TEA-21
Hiawatha (Minneapolis, MN)	1999	1999	2001	2001	2004	2020	2007	NA	NA	TEA-21
Med Center Ext (Memphis, TN)	1998	1998	2000	2001	2004	2020	2007	ISTEA	ISTEA	TEA-21
Tren Urbano (San Juan, PR)	1993	1993	1996	1999	2005	2010	2007	ISTEA	ISTEA	ISTEA
Reconstruction II (Pittsburgh, PA)	1994	1994	2001	2001	2005	2010	2007	ISTEA	ISTEA	TEA-21
Mission Valley East (San Diego, CA)	1998	1998	2000	2000	2005	2015	2007	ISTEA	ISTEA	TEA-21
Douglas Reconstr. (Chicago, IL)	1999	1999	2001	2001	2005	2020	2007	TEA-21	TEA-21	TEA-21
Elizabeth Link (Newark, NJ)	1993	1993	2000	2000	2006	2015	2008	ISTEA	ISTEA	TEA-21
Hudson Bergen I (Jersey City, NJ)	NA	1992	1996	1996	2002	2010	2008	NA	ISTEA	ISTEA
Hudson Bergen II (Jersey City, NJ)	NA	1992	2000	2000	2006	2010	2008	NA	ISTEA	TEA-21
North Central (Chicago, IL)	1998	1998	2001	2001	2006	2020	2006	ISTEA	ISTEA	TEA-21
Southwest (Chicago, IL)	1998	1998	2001	2001	2006	2020	2006	ISTEA	ISTEA	TEA-21
UP West Ext (Chicago, IL)	1998	1998	2001	2001	2006	2020	2006	ISTEA	ISTEA	TEA-21

Table B-1: Time variables for study sample (3/3)

	Year of …							Funding legislation		
Project name	Initial cost forecast	Ridership forecast	Final cost forecast	Construction start	Project opening	Ridership forecast horizon	Observed ridership	Initial cost forecast	Ridership forecast	Final cost forecast
Southeast Corridor (Denver, CO)	1998	1998	2000	2000	2006	2020	2007	ISTEA	ISTEA	TEA-21
Central Double Track (Baltimore)	1999	1999	2001	2001	2006	2020	2008	TEA-21	TEA-21	TEA-21
Ft Lauderdale Dbl-Track (Miami)	1999	1999	2000	2000	2007	2015	2007	TEA-21	TEA-21	TEA-21
South Corridor (Charlotte, NC)	1999	1998	2005	2005	2007	2025	2009	TEA-21	TEA-21	TEA-21
Sprinter light rail (Oceanside, CA)	1995	1995	2002	2002	2008	2015	2010	ISTEA	ISTEA	TEA-21
Euclid Corridor (Cleveland, OH)	1997	1997	2004	2004	2008	2010	2010	ISTEA	ISTEA	TEA-21
East Valley (Phoenix, AZ)	1998	1998	2005	2005	2008	2020	2011	ISTEA	ISTEA	TEA-21
Frontrunner N (Salt Lake City, UT)	2003	2003	2006	2006	2008	2008	2009	TEA-21	TEA-21	SAFETEA
Gold Line Ext. (Los Angeles, CA)	2000	2000	2004	2004	2009	2020	2011	TEA-21	TEA-21	TEA-21
Central, Airport Links (Seattle)	2000	2000	2001	2003	2009	2009	2011	TEA-21	TEA-21	TEA-21
Westside Express (Portland, OR)	2001	2001	2006	2006	2009	2009	2009	TEA-21	TEA-21	SAFETEA
Green Line, (Portland, OR)	2004	2004	2007	2007	2009	2009	2011	TEA-21	TEA-21	SAFETEA
Northstar (Minneapolis, MN)	2005	2005	2007	2007	2009	2009	2011	TEA-21	TEA-21	SAFETEA
NW-SE Line (Dallas, TX)	2001	2001	2006	2006	2010	2010	2012	TEA-21	TEA-21	SAFETEA
The Tide (Norfolk, VA)	2002	2002	2007	2008	2011	2011	2013	TEA-21	TEA-21	SAFETEA
Mountainlink (Flagstaff, AZ)	2009	2009	2011	2011	2011	2011	2013	SAFETEA	SAFETEA	SAFETEA

Table B-2: Sources of time variables for study sample (1/3)

	Year of…						
Project name	Initial cost forecast	Ridership forecast	Final cost forecast	Construction start	Project opening	Ridership forecast horizon	Observed ridership
Heavy Rail (Baltimore, MD)	(1)	(1)	NA	(2)	(1)	(1)	(1)
Heavy Rail (Washington, DC)	(1)	(1)	NA	(3)	(1)	(1)	(1)
Heavy Rail (Miami, FL)	(1)	(1)	NA	(4)	(1)	(1)	(1)
Heavy Rail (Atlanta, GA)	(1)	NA	NA	(5)	(1)	NA	(1)
Light Rail (Buffalo, NY)	(1)	(1)	NA	(6)	(1)	(1)	(1)
People Mover (Miami, FL)	(1)	(1)	NA	(7)	(1)	(1)	(1)
Benfield Transitway (Portland, OR)	(1)	(1)	NA	(8)	(1)	(1)	(1)
Light Rail (Pittsburgh, PA)	(1)	(1)	NA	(9)	(1)	(1)	(1)
People Mover (Detroit, MI)	(1)	(1)	NA	(10)	(1)	(1)	(1)
Light Rail (Sacramento, CA)	(1)	(1)	NA	(11)	(1)	(1)	(1)
El Cajon Extension (San Diego, CA)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Downtown Transit (Seattle, WA)	(7)	NA	(7)	(7)	(7)	NA	(7)
Guadalupe (San Jose, CA)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Orange Line (Chicago, IL)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
SW Transitway (Houston, TX)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Metrolink (St Louis, MO)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Metromover Extensions (Miami, FL)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
I-25 Transitway (Denver, CO)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Metro Extension (Baltimore, MD)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Colma station (San Francisco, CA)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
South Oak Cliff (Dallas, TX)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
LRT Extensions (Baltimore, MD)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Hillsboro Extension (Portland, OR)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
N/S LRT (Salt Lake City, UT)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Tasman West LRT (San Jose, CA)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Skyway Express (Jacksonville, FL)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Red Line (Los Angeles, CA)	NA	(7)	NA	(7)	(7)	(7)	(7)

Table B-2: Sources of time variables in study sample (2/3)

	Year of …						
Project name	**Initial cost forecast**	**Ridership forecast**	**Final cost forecast**	**Construction start**	**Project opening**	**Ridership forecast horizon**	**Observed ridership**
Red Line MOS 1 (Los Angeles, CA)	(7)	NA	(7)	(7)	(7)	NA	NA
Red Line MOS 2 (Los Angeles, CA)	(7)	NA	(7)	(7)	(7)	NA	NA
Red Line MOS 3 (Los Angeles, CA)	(7)	NA	(7)	(7)	(7)	NA	NA
North Line Extension (Atlanta, GA)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Airport Busway (Pittsburgh, PA)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
Southwest Light Rail (Denver, CO)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
University Ext (Salt Lake City, UT)	(14)	(14)	(14)	(14)	(14)	(14)	(14)
St Clair Extension (St Louis, MO)	(7)	(7)	(7)	(7)	(7)	(7)	(7)
North Central Ext (Dallas, TX)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
BART to SFO (San Francisco, CA)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Med Center Ext (Salt Lake City, UT)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
South LRT (Sacramento, CA)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Piers Transitway (Boston, MA)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Largo Extension (Washington, DC)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Interstate MAX (Portland, OR)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Hiawatha (Minneapolis, MN)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Med Center Ext (Memphis, TN)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Tren Urbano (San Juan, PR)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Reconstruction II (Pittsburgh, PA)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Mission Valley East (San Diego, CA)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Douglas Reconstr. (Chicago, IL)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Elizabeth Link (Newark, NJ)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Hudson Bergen I (Jersey City, NJ)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Hudson Bergen II (Jersey City, NJ)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
North Central (Chicago, IL)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Southwest (Chicago, IL)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
UP West Ext (Chicago, IL)	(12)	(12)	(12)	(12)	(12)	(12)	(12)

Table B-2: Sources of time variables in study sample (3/3)

	Year of…						
Project name	Initial cost forecast	Ridership forecast	Final cost forecast	Construction start	Project opening	Ridership forecast horizon	Observed ridership
Southeast Corridor (Denver, CO)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Central Double Track (Baltimore)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
Ft Lauderdale Dbl-Track (Miami)	(12)	(12)	(12)	(12)	(12)	(12)	(12)
South Corridor (Charlotte, NC)	(13)	(13)	(13)	(13)	(13)	(13)	(13)
Sprinter light rail (Oceanside, CA)	(13)	(13)	(13)	(13)	(13)	(13)	(13)
Euclid Corridor (Cleveland, OH)	(14)	(14)	(14)	(14)	(14)	(14)	(14)
East Valley (Phoenix, AZ)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
Frontrunner N (Salt Lake City, UT)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
Gold Line Ext. (Los Angeles, CA)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
Central, Airport Links (Seattle)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
Westside Express (Portland, OR)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
Green Line, (Portland, OR)	(16)	(16)	(16)	(16)	(16)	(16)	(16)
Northstar (Minneapolis, MN)	(15)	(15)	(15)	(15)	(15)	(15)	(15)
NW-SE Line (Dallas, TX)	(16)	(16)	(16)	(16)	(16)	(16)	(16)
The Tide (Norfolk, VA)	(17)	(17)	(17)	(17)	(17)	(17)	(17)
Mountainlink (Flagstaff, AZ)	(16)	(16)	(16)	(16)	(16)	(16)	(16)

Table B-3: Accuracy variables for study sample (1/3)

	Cost (millions)			Ridership (average weekday boardings)	
Project name	Initial Forecast	Final Forecast	Actual	Forecast	Actual
Heavy Rail (Baltimore, MD)	$405	NA	$790	103,000	42,600
Heavy Rail (Washington, DC)	$1,713	NA	$4,375	569,600	411,600
Heavy Rail (Miami, FL)	$795	NA	$1,042	239,900	35,400
Heavy Rail (Atlanta, GA)	$793	NA	$1,838	NA	184,500
Light Rail (Buffalo, NY)	$336	NA	$536	92,000	29,200
People Mover (Miami, FL)	$77	NA	$153	41,000	10,800
Benfield Transitway (Portland, OR)	$188	NA	$240	42,500	19,700
Light Rail (Pittsburgh, PA)	$548	NA	$537	90,500	30,600
People Mover (Detroit, MI)	$109	NA	$197	67,700	41,000
Light Rail (Sacramento, CA)	$147	NA	$172	50,000	14,400
El Cajon Extension (San Diego, CA)	$114	$100	$103	21,600	23,478
Downtown Transit (Seattle, WA)	$300	$400	$469	NA	27,800
Guadalupe (San Jose, CA)	$258	$395	$380	41,200	19,738
Orange Line (Chicago, IL)	$604	$438	$522	118,760	54,042
SW Transitway (Houston, TX)	$99	$104	$98	27,280	8,875
Metrolink (St Louis, MO)	$380	$456	$464	41,800	37,045
Metromover Extensions (Miami, FL)	$221	$161	$228	20,404	4,209
I-25 Transitway (Denver, CO)	$190	$205	$228	NA	9,959
Metro Extension (Baltimore, MD)	$314	$311	$353	13,600	10,128
Colma station (San Francisco, CA)	$113	$172	$180	15,200	13,482
South Oak Cliff (Dallas, TX)	$325	$377	$360	34,800	26,884
LRT Extensions (Baltimore, MD)	$82	$110	$116	11,804	8,272
Hillsboro Extension (Portland, OR)	$559	$887	$964	40,514	43,876
N/S LRT (Salt Lake City, UT)	$306	$300	$299	26,500	22,100
Tasman West LRT (San Jose, CA)	$466	$346	$325	14,875	8,244
Skyway Express (Jacksonville, FL)	$86	$142	$137	42,472	2,054
Red Line (Los Angeles, CA)	NA	NA	NA	297,733	128,659
Red Line MOS 1 (Los Angeles, CA)	NA	$947	$1,439	NA	NA

Table B-3: Accuracy variables for study sample (2/3)

	Cost (millions)			Ridership (average weekday boardings)	
Project name	**Initial Forecast**	**Final Forecast**	**Actual**	**Forecast**	**Actual**
Red Line MOS 2 (Los Angeles, CA)	NA	$1,290	$1,718	NA	NA
Red Line MOS 3 (Los Angeles, CA)	NA	$1,269	$1,313	NA	NA
North Line Extension (Atlanta, GA)	$440	$352	$473	57,120	22,328
Airport Busway (Pittsburgh, PA)	$274	$362	$322	23,369	9,000
Southwest Light Rail (Denver, CO)	$150	$177	$178	22,000	19,083
University Ext (Salt Lake City, UT)	NA	$114	$108	7,577	11,359
St Clair Extension (St Louis, MO)	$368	$322	$339	11,960	15,976
North Central Ext (Dallas, TX)	$333	$461	$437	11,000	13,581
BART to SFO (San Francisco, CA)	$1,194	$1,186	$1,552	67,400	26,284
Med Center Ext (Salt Lake City, UT)	NA	$91	$85	2,473	2,640
South LRT (Sacramento, CA)	$202	$220	$219	12,550	8,734
Piers Transitway (Boston, MA)	$398	$457	$600	29,700	12,500
Largo Extension (Washington, DC)	$375	$413	$426	14,270	6,361
Interstate MAX (Portland, OR)	$265	$322	$324	17,030	12,785
Hiawatha (Minneapolis, MN)	$541	$675	$697	24,600	27,871
Med Center Ext (Memphis, TN)	$36	$73	$58	4,200	720
Tren Urbano (San Juan, PR)	$1,086	$1,281	$2,228	113,643	27,567
Reconstruction II (Pittsburgh, PA)	$401	$363	$385	49,000	23,411
Mission Valley East (San Diego, CA)	$387	$427	$506	10,795	7,572
Douglas Reconstr. (Chicago, IL)	$451	$473	$441	33,000	25,106
Elizabeth Link (Newark, NJ)	$181	$215	$208	12,500	2,000
Hudson Bergen I (Jersey City, NJ)	NA	$844	$860	31,300	20,868
Hudson Bergen II (Jersey City, NJ)	NA	$998	$886	34,860	17,322
North Central (Chicago, IL)	$205	$225	$217	8,400	5,338
Southwest (Chicago, IL)	$179	$191	$185	13,800	8,811
UP West Ext (Chicago, IL)	$99	$128	$106	3,900	2,078
Southeast Corridor (Denver, CO)	$585	$868	$851	30,000	26,192
Central Double Track (Baltimore)	$151	$154	$152	44,000	26,987

Table B-3: Accuracy variables for study sample (3/3)

	Cost (millions)			Ridership (average weekday boardings)	
Project name	Initial Forecast	Final Forecast	Actual	Forecast	Actual
Ft Lauderdale Dbl-Track (Miami)	$330	$339	$346	42,100	11,503
South Corridor (Charlotte, NC)	$331	$427	$463	14,000	14,370
Sprinter light rail (Oceanside, CA)	$214	$352	$478	15,100	7,569
Euclid Corridor (Cleveland, OH)	$185	$168	$168	21,100	14,300
East Valley (Phoenix, AZ)	$1,076	$1,412	$1,405	25,800	40,700
Frontrunner N (Salt Lake City, UT)	$408	$612	$614	8,400	4,700
Gold Line Ext. (Los Angeles, CA)	$760	$899	$899	18,000	13,000
Central, Airport Links (Seattle)	$1,858	$2,668	$2,558	34,900	23,400
Westside Express (Portland, OR)	$85	$117	$162	2,400	1,200
Green Line, (Portland, OR)	$495	$559	$576	30,400	24,000
Northstar (Minneapolis, MN)	$265	$320	$309	4,100	2,200
NW-SE Line (Dallas, TX)	$1,151	$1,406	$1,406	40,300	31,000
The Tide (Norfolk, VA)	$195	$232	$315	2,900	4,600
Mountainlink (Flagstaff, AZ)	$10	$8	$8	4,150	4,200

Table B-4: Sources of accuracy variables for study sample (1/3)

	Cost (millions)			Ridership (average weekday boardings)	
Project name	Initial Forecast	Final Forecast	Actual	Forecast	Actual
Heavy Rail (Baltimore, MD)	(1)	NA	(1)	(1)	(1)
Heavy Rail (Washington, DC)	(1)	NA	(1)	(1)	(1)
Heavy Rail (Miami, FL)	(1)	NA	(1)	(1)	(1)
Heavy Rail (Atlanta, GA)	(1)	NA	(1)	NA	(1)
Light Rail (Buffalo, NY)	(1)	NA	(1)	(1)	(1)
People Mover (Miami, FL)	(1)	NA	(1)	(1)	(1)
Benfield Transitway (Portland, OR)	(1)	NA	(1)	(1)	(1)
Light Rail (Pittsburgh, PA)	(18)	NA	(1)	(1)	(1)
People Mover (Detroit, MI)	(1)	NA	(1)	(1)	(1)
Light Rail (Sacramento, CA)	(1)	NA	(1)	(1)	(1)
El Cajon Extension (San Diego, CA)	(7)	(7)	(7)	(7)	(7)
Downtown Transit (Seattle, WA)	(7)	(7)	(7)	NA	(7)
Guadalupe (San Jose, CA)	(7)	(7)	(7)	(7)	(7)
Orange Line (Chicago, IL)	(7)	(7)	(7)	(7)	(7)
SW Transitway (Houston, TX)	(7)	(7)	(7)	(7)	(7)
Metrolink (St Louis, MO)	(7)	(7)	(7)	(7)	(7)
Metromover Extensions (Miami, FL)	(7)	(7)	(7)	(7)	(7)
I-25 Transitway (Denver, CO)	(7)	(7)	(7)	NA	(7)
Metro Extension (Baltimore, MD)	(7)	(7)	(7)	(7)	(7)
Colma station (San Francisco, CA)	(7)	(7)	(7)	(7)	(7)
South Oak Cliff (Dallas, TX)	(7)	(7)	(7)	(7)	(7)
LRT Extensions (Baltimore, MD)	(7)	(7)	(7)	(7)	(7)
Hillsboro Extension (Portland, OR)	(7)	(7)	(7)	(7)	(7)
N/S LRT (Salt Lake City, UT)	(7)	(7)	(7)	(7)	(7)
Tasman West LRT (San Jose, CA)	(7)	(7)	(7)	(7)	(7)
Skyway Express (Jacksonville, FL)	(7)	(7)	(7)	(7)	(7)
Red Line (Los Angeles, CA)	NA	NA	NA	(7)	(7)
Red Line MOS 1 (Los Angeles, CA)	NA	(7)	(7)	NA	NA

Table B-4: Sources of accuracy variables for study sample (2/3)

	Cost (millions)			Ridership (average weekday boardings)	
Project name	**Initial Forecast**	**Final Forecast**	**Actual**	**Forecast**	**Actual**
Red Line MOS 2 (Los Angeles, CA)	NA	(7)	(7)	NA	NA
Red Line MOS 3 (Los Angeles, CA)	NA	(7)	(7)	NA	NA
North Line Extension (Atlanta, GA)	(7)	(7)	(7)	(7)	(7)
Airport Busway (Pittsburgh, PA)	(7)	(7)	(7)	(7)	(7)
Southwest Light Rail (Denver, CO)	(7)	(7)	(7)	(7)	(7)
University Ext (Salt Lake City, UT)	NA	(12)	(12)	(12)	(12)
St Clair Extension (St Louis, MO)	(7)	(7)	(7)	(7)	(7)
North Central Ext (Dallas, TX)	(12)	(12)	(12)	(12)	(12)
BART to SFO (San Francisco, CA)	(12)	(12)	(12)	(12)	(12)
Med Center Ext (Salt Lake City, UT)	NA	(12)	(12)	(12)	(12)
South LRT (Sacramento, CA)	(12)	(12)	(12)	(12)	(12)
Piers Transitway (Boston, MA)	(12)	(12)	(12)	(12)	(12)
Largo Extension (Washington, DC)	(12)	(12)	(12)	(12)	(12)
Interstate MAX (Portland, OR)	(12)	(12)	(12)	(12)	(12)
Hiawatha (Minneapolis, MN)	NA	(12)	(12)	(12)	(12)
Med Center Ext (Memphis, TN)	(12)	(12)	(12)	(12)	(12)
Tren Urbano (San Juan, PR)	(12)	(12)	(12)	(12)	(12)
Reconstruction II (Pittsburgh, PA)	(12)	(12)	(12)	(12)	(12)
Mission Valley East (San Diego, CA)	(12)	(12)	(12)	(12)	(12)
Douglas Reconstr. (Chicago, IL)	(12)	(12)	(12)	(12)	(12)
Elizabeth Link (Newark, NJ)	(12)	(12)	(12)	(12)	(12)
Hudson Bergen I (Jersey City, NJ)	NA	(19)	(12)	(12)	(12)
Hudson Bergen II (Jersey City, NJ)	NA	(19)	(12)	(12)	(12)
North Central (Chicago, IL)	(12)	(12)	(12)	(12)	(12)
Southwest (Chicago, IL)	(12)	(12)	(12)	(12)	(12)
UP West Ext (Chicago, IL)	(12)	(12)	(12)	(12)	(12)
Southeast Corridor (Denver, CO)	(12)	(12)	(12)	(12)	(12)
Central Double Track (Baltimore)	(12)	(12)	(12)	(12)	(12)

Table B-4: Sources of accuracy variables for study sample (3/3)

	Cost (millions)			Ridership (average weekday boardings)	
Project name	**Initial Forecast**	**Final Forecast**	**Actual**	**Forecast**	**Actual**
Ft Lauderdale Dbl-Track (Miami)	(12)	(12)	(12)	(12)	(12)
South Corridor (Charlotte, NC)	(13)	(13)	(13)	(13)	(13)
Sprinter light rail (Oceanside, CA)	(13)	(13)	(13)	(13)	(13)
Euclid Corridor (Cleveland, OH)	(14)	(14)	(14)	(14)	(14)
East Valley (Phoenix, AZ)	(15)	(15)	(15)	(15)	(15)
Frontrunner N (Salt Lake City, UT)	(15)	(15)	(15)	(15)	(15)
Gold Line Ext. (Los Angeles, CA)	(15)	(15)	(15)	(15)	(15)
Central, Airport Links (Seattle)	(15)	(15)	(15)	(15)	(15)
Westside Express (Portland, OR)	(15)	(15)	(15)	(15)	(15)
Green Line, (Portland, OR)	(16)	(16)	(16)	(16)	(16)
Northstar (Minneapolis, MN)	(15)	(15)	(15)	(15)	(15)
NW-SE Line (Dallas, TX)	(16)	(16)	(16)	(16)	(16)
The Tide (Norfolk, VA)	(17)	(17)	(17)	(17)	(17)
Mountainlink (Flagstaff, AZ)	(16)	(16)	(16)	(16)	(16)

Table B-5: Funding, scale, and experience variables for study sample (1/3)

Project name	Federal funding share	Project sequence	Project length (miles)	Transit mode	System mileage at time of…		
					Initial cost forecast	Ridership forecast	Final cost forecast
Heavy Rail (Baltimore, MD)	79%	Initial line	7.6	Heavy rail	0	0	NA
Heavy Rail (Washington, DC)	67%	Initial line	60.5	Heavy rail	0	0	NA
Heavy Rail (Miami, FL)	69%	Initial line	21.0	Heavy rail	0	0	NA
Heavy Rail (Atlanta, GA)	67%	Initial line	26.8	Heavy rail	0	0	NA
Light Rail (Buffalo, NY)	79%	Initial line	6.4	Light rail	0	0	NA
People Mover (Miami, FL)	53%	Initial line	2.0	People mover	0	0	NA
Benfield Transitway (Portland, OR)	83%	Initial line	2.9	Light rail	0	0	NA
Light Rail (Pittsburgh, PA)	80%	Initial line	15.1	Light rail	0	0	NA
People Mover (Detroit, MI)	79%	Initial line	10.5	People mover	0	0	NA
Light Rail (Sacramento, CA)	59%	Initial line	18.3	Light rail	0	0	NA
El Cajon Extension (San Diego, CA)	61%	Expansion	11.1	Light rail	15.9	15.9	15.9
Downtown Transit (Seattle, WA)	42%	Renovation	1.3	Transitway	NA	NA	NA
Guadalupe (San Jose, CA)	68%	Initial line	20.0	Light rail	0	0	0
Orange Line (Chicago, IL)	67%	Expansion	9.0	Heavy rail	112	112	112
SW Transitway (Houston, TX)	63%	Renovation	9.7	Transitway	NA	NA	NA
Metrolink (St Louis, MO)	75%	Initial line	18.0	Light rail	0	0	0
Metromover Extensions (Miami, FL)	59%	Expansion	2.5	People mover	1.9	1.9	1.9
I-25 Transitway (Denver, CO)	44%	Renovation	6.6	Transitway	NA	NA	NA
Metro Extension (Baltimore, MD)	78%	Expansion	1.5	Heavy rail	14	14	14
Colma station (San Francisco, CA)	60%	Expansion	0.9	Heavy rail	98.3	98.3	98.3
South Oak Cliff (Dallas, TX)	44%	Initial line	9.6	Light rail	0	0	0
LRT Extensions (Baltimore, MD)	73%	Expansion	7.3	Light rail	22	22	22
Hillsboro Extension (Portland, OR)	73%	Expansion	17.1	Light rail	15.1	15.1	15.1
N/S LRT (Salt Lake City, UT)	82%	Initial line	15	Light rail	0	0	0
Tasman West LRT (San Jose, CA)	75%	Initial line	7.6	Light rail	0	0	0
Skyway Express (Jacksonville, FL)	50%	Initial line	2.5	People mover	0	0	0
Red Line (Los Angeles, CA)	NA	Initial line	17.0	Heavy rail	NA	0	NA
Red Line MOS 1 (Los Angeles, CA)	48%	Initial line	4.4	Heavy rail	NA	NA	0

Table B-5: Funding, scale, and experience variables for study sample (2/3)

Project name	Federal funding share	Project sequence	Project length (miles)	Transit mode	System mileage at time of…		
					Initial cost forecast	Ridership forecast	Final cost forecast
Red Line MOS 2 (Los Angeles, CA)	39%	Expansion	6.7	Heavy rail	NA	NA	0
Red Line MOS 3 (Los Angeles, CA)	71%	Expansion	5.9	Heavy rail	NA	NA	4.4
North Line Extension (Atlanta, GA)	78%	Expansion	3.1	Heavy rail	26.8	26.8	26.8
Airport Busway (Pittsburgh, PA)	80%	Renovation	6.1	Transitway	NA	NA	NA
Southwest Light Rail (Denver, CO)	68%	Expansion	8.7	Light rail	0	0	5.3
University Ext (Salt Lake City, UT)	90%	Expansion	2.5	Light rail	NA	NA	15
St Clair Extension (St Louis, MO)	72%	Expansion	17.0	Light rail	0	0	18
North Central Ext (Dallas, TX)	76%	Expansion	12.5	Light rail	0	0	20
BART to SFO (San Francisco, CA)	48%	Expansion	8.7	Heavy rail	98.3	98.3	99.2
Med Center Ext (Salt Lake City, UT)	63%	Expansion	1.5	Light rail	NA	NA	17.5
South LRT (Sacramento, CA)	52%	Expansion	6.3	Light rail	18.3	18.3	18.3
Piers Transitway (Boston, MA)	55%	Renovation	1.0	Transitway	NA	NA	NA
Largo Extension (Washington, DC)	85%	Expansion	3.1	Heavy rail	89.0	89.0	96.6
Interstate MAX (Portland, OR)	87%	Expansion	5.8	Light rail	15.1	15.1	32.2
Hiawatha (Minneapolis, MN)	55%	Initial line	11.6	Light rail	0	0	0
Med Center Ext (Memphis, TN)	103%	Expansion	2.0	Light rail	5.0	5.0	5.0
Tren Urbano (San Juan, PR)	32%	Initial line	10.7	Heavy rail	0	0	0
Reconstruction II (Pittsburgh, PA)	61%	Expansion	13	Light rail	12.0	12.0	12.0
Mission Valley East (San Diego, CA)	68%	Expansion	5.9	Light rail	270	27.0	27.0
Douglas Reconstr. (Chicago, IL)	87%	Renovation	6.6	Heavy rail	NA	NA	NA
Elizabeth Link (Newark, NJ)	81%	Initial line	1.0	Light rail	0	0	0
Hudson Bergen I (Jersey City, NJ)	69%	Initial line	9.3	Light rail	0	0	0
Hudson Bergen II (Jersey City, NJ)	74%	Expansion	6.1	Light rail	0	0	0
North Central (Chicago, IL)	62%	Renovation	6.3	Commuter rail	NA	NA	NA
Southwest (Chicago, IL)	56%	Expansion	11	Commuter rail	468	468	468
UP West Ext (Chicago, IL)	88%	Expansion	8.5	Commuter rail	468	468	468
Southeast Corridor (Denver, CO)	62%	Expansion	19.1	Light rail	5.3	5.3	5.3
Central Double Track (Baltimore)	81%	Renovation	9.4	Light rail	NA	NA	NA

Table B-5: Funding, scale, and experience variables for study sample (3/3)

Project name	Federal funding share	Project sequence	Project length (miles)	Transit mode	System mileage at time of...		
					Initial cost forecast	Ridership forecast	Final cost forecast
Ft Lauderdale Dbl-Track (Miami)	58%	Renovation	43.6	Commuter rail	NA	NA	NA
Blue line (Charlotte, NC)	43%	Initial line	9.6	Light rail	0	0	0
Sprinter light rail (Oceanside, CA)	32%	Initial line	22.0	Light rail	0	0	0
Euclid Corridor (Cleveland, OH)	85%	Initial line	7.1	Rapid bus	0	0	0
East Valley (Phoenix, AZ)	46%	Initial line	19.7	Light rail	0	0	0
Frontrunner N (Salt Lake City, UT)	80%	Initial line	44.0	Commuter rail	0	0	0
Gold Line Ext. (Los Angeles, CA)	78%	Expansion	6.0	Light rail	42.0	42.0	55.7
Central, Airport Links (Seattle)	20%	Initial line	15.6	Light rail	0	0	0
Westside Express (Portland, OR)	43%	Initial line	14.7	Commuter rail	0	0	0
Green Line, (Portland, OR)	76%	Expansion	14.5	Light rail	32.2	32.2	38.0
Northstar (Minneapolis, MN)	53%	Initial line	40	Commuter rail	0	0	0
NW-SE Line (Dallas, TX)	50%	Expansion	20.9	Light rail	20	20	44.7
The Tide (Norfolk, VA)	53%	Initial line	7.3	Light rail	0	0	0
Mountainlink (Flagstaff, AZ)	76%	Initial line	4.0	Rapid bus	0	0	0

Table B-6: Sources of funding, scale, and experience variables for study sample (1/3)

Project name	Federal funding share	Project sequence	Project length (miles)	Transit mode	System mileage at time of… Initial cost forecast	Ridership forecast	Final cost forecast
Heavy Rail (Baltimore, MD)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Heavy Rail (Washington, DC)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Heavy Rail (Miami, FL)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Heavy Rail (Atlanta, GA)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Light Rail (Buffalo, NY)	(1)	(1)	(1)	(1)	(1)	(1)	NA
People Mover (Miami, FL)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Benfield Transitway (Portland, OR)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Light Rail (Pittsburgh, PA)	(1)	(1)	(1)	(1)	(1)	(1)	NA
People Mover (Detroit, MI)	(1)	(1)	(1)	(1)	(1)	(1)	NA
Light Rail (Sacramento, CA)	(1)	(1)	(1)	(1)	(1)	(1)	NA
El Cajon Extension (San Diego, CA)	(20)	(7)	(7)	(7)	(21)	(21)	(21)
Downtown Transit (Seattle, WA)	(20)	(7)	(7)	(7)	NA	NA	NA
Guadalupe (San Jose, CA)	(20)	(7)	(7)	(7)	(7)	(7)	(7)
Orange Line (Chicago, IL)	(20)	(7)	(7)	(7)	(22)	(22)	(22)
SW Transitway (Houston, TX)	(20)	(7)	(7)	(7)	NA	NA	NA
Metrolink (St Louis, MO)	(23)	(7)	(7)	(7)	(7)	(7)	(7)
Metromover Extensions (Miami, FL)	(24)	(7)	(7)	(7)	(7)	(7)	(7)
I-25 Transitway (Denver, CO)	(24)	(7)	(7)	(7)	NA	NA	NA
Metro Extension (Baltimore, MD)	(20)	(7)	(7)	(7)	(2)	(2)	(2)
Colma station (San Francisco, CA)	(25)	(7)	(7)	(7)	(22)	(22)	(22)
South Oak Cliff (Dallas, TX)	(26)	(7)	(7)	(7)	(7)	(7)	(7)
LRT Extensions (Baltimore, MD)	(26)	(7)	(7)	(7)	(7)	(7)	(7)
Hillsboro Extension (Portland, OR)	(26)	(7)	(7)	(7)	(1)	(1)	(1)
N/S LRT (Salt Lake City, UT)	(27)	(7)	(7)	(7)	(7)	(7)	(7)
Tasman West LRT (San Jose, CA)	(27)	(7)	(7)	(7)	(7)	(7)	(7)
Skyway Express (Jacksonville, FL)	(28)	(7)	(7)	(7)	(7)	(7)	(7)
Red Line (Los Angeles, CA)	NA	(7)	(7)	(7)	NA	(7)	NA
Red Line MOS 1 (Los Angeles, CA)	(20)	(7)	(7)	(7)	NA	NA	(7)

Table B-6: Sources of funding, scale, and experience variables for study sample (2/3)

Project name	Federal funding share	Project sequence	Project length (miles)	Transit mode	System mileage at time of…		
					Initial cost forecast	Ridership forecast	Final cost forecast
Red Line MOS 2 (Los Angeles, CA)	(25)	(7)	(7)	(7)	NA	NA	(7)
Red Line MOS 3 (Los Angeles, CA)	(27)	(7)	(7)	(7)	NA	NA	(7)
North Line Extension (Atlanta, GA)	(29)	(7)	(7)	(7)	(1)	(1)	(1)
Airport Busway (Pittsburgh, PA)	(28)	(7)	(7)	(7)	NA	NA	NA
Southwest Light Rail (Denver, CO)	(27)	(7)	(7)	(7)	(30)	(30)	(30)
University Ext (Salt Lake City, UT)	(27)	(12)	(12)	(12)	(7)	(7)	(7)
St Clair Extension (St Louis, MO)	(29)	(7)	(7)	(7)	(7)	(7)	(7)
North Central Ext (Dallas, TX)	(29)	(12)	(12)	(12)	(12)	(12)	(12)
BART to SFO (San Francisco, CA)	(31)	(12)	(12)	(12)	(12)	(12)	(12)
Med Center Ext (Salt Lake City, UT)	(32)	(12)	(12)	(12)	NA	NA	(25)
South LRT (Sacramento, CA)	(27)	(12)	(12)	(12)	(1)	(1)	(1)
Piers Transitway (Boston, MA)	(33)	(12)	(12)	(12)	NA	NA	NA
Largo Extension (Washington, DC)	(34)	(12)	(12)	(12)	(35)	(35)	(35)
Interstate MAX (Portland, OR)	(31)	(12)	(12)	(12)	(1)	(1)	(1, 7)
Hiawatha (Minneapolis, MN)	(27)	(12)	(12)	(12)	(12)	(12)	(12)
Med Center Ext (Memphis, TN)	(27)	(12)	(12)	(12)	(36)	(36)	(36)
Tren Urbano (San Juan, PR)	(31)	(12)	(12)	(12)	(12)	(12)	(12)
Reconstruction II (Pittsburgh, PA)	(27)	(12)	(12)	(12)	(12)	(12)	(12)
Mission Valley East (San Diego, CA)	(31)	(12)	(12)	(12)	(7, 21)	(7, 21)	(7, 21)
Douglas Reconstr. (Chicago, IL)	(31)	(12)	(12)	(12)	NA	NA	NA
Elizabeth Link (Newark, NJ)	(27)	(12)	(12)	(12)	(12)	(12)	(12)
Hudson Bergen I (Jersey City, NJ)	(32)	(12)	(12)	(12)	(12)	(12)	(12)
Hudson Bergen II (Jersey City, NJ)	(31)	(12)	(12)	(12)	(12)	(12)	(12)
North Central (Chicago, IL)	(37)	(12)	(12)	(12)	NA	NA	NA
Southwest (Chicago, IL)	(37)	(12)	(12)	(12)	(38)	(38)	(38)
UP West Ext (Chicago, IL)	(37)	(12)	(12)	(12)	(38)	(38)	(38)
Southeast Corridor (Denver, CO)	(31)	(12)	(12)	(12)	(30)	(30)	(30)
Central Double Track (Baltimore)	(31)	(12)	(12)	(12)	NA	NA	NA

Table B-6: Sources of funding, scale, and experience variables for study sample (3/3)

Project name	Federal funding share	Project sequence	Project length (miles)	Transit mode	System mileage at time of…		
					Initial cost forecast	Ridership forecast	Final cost forecast
Ft Lauderdale Dbl-Track (Miami)	(33)	(12)	(12)	(12)	NA	NA	NA
South Corridor (Charlotte, NC)	(31)	(13)	(13)	(13)	(13)	(13)	(13)
Sprinter light rail (Oceanside, CA)	(31)	(13)	(13)	(13)	(13)	(13)	(13)
Euclid Corridor (Cleveland, OH)	(31)	(14)	(14)	(14)	(14)	(14)	(14)
East Valley (Phoenix, AZ)	(34)	(15)	(15)	(15)	(15)	(15)	(15)
Frontrunner N (Salt Lake City, UT)	(34)	(15)	(15)	(15)	(15)	(15)	(15)
Gold Line Ext. (Los Angeles, CA)	(34)	(15)	(15)	(15)	(39)	(39)	(39)
Central, Airport Links (Seattle)	(34)	(15)	(15)	(15)	(15)	(15)	(15)
Westside Express (Portland, OR)	(31)	(15)	(15)	(15)	(15)	(15)	(15)
Green Line, (Portland, OR)	(34)	(16)	(16)	(16)	(1, 7)	(1, 7)	(1, 7)
Northstar (Minneapolis, MN)	(34)	(15)	(15)	(15)	(15)	(15)	(15)
NW-SE Line (Dallas, TX)	(34)	(16)	(16)	(16)	(40)	(40)	(40)
The Tide (Norfolk, VA)	(34)	(17)	(17)	(17)	(17)	(17)	(17)
Mountainlink (Flagstaff, AZ)	(34)	(16)	(16)	(16)	(16)	(16)	(16)

Table B-7: External control variables for study sample (1/3)

Project name	State average price per gallon of gasoline at time of…					County population (thousands) at time of…					State unemployment rate (June, seasonally adjusted) at time of…				
	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership
Heavy Rail (Baltimore, MD)	0.36	0.54	NA	1.13	0.97	900	887	NA	770	751	NA	NA	NA	7.2	4.3
Heavy Rail (Washington, DC)	NA	NA	NA	1.24	0.99	762	762	NA	635	638	NA	NA	NA	8.4	7.7
Heavy Rail (Miami, FL)	0.62	0.57	NA	1.09	0.86	1,554	1,518	NA	1,781	1,875	6.7	8.3	NA	6.1	5.0
Heavy Rail (Atlanta, GA)	0.35	NA	NA	0.76	0.80	604	NA	NA	625	631	NA	NA	NA	6.1	5.4
Light Rail (Buffalo, NY)	0.63	0.63	NA	0.79	0.94	1,045	1,045	NA	987	973	9.1	9.1	NA	6.5	4.9
People Mover (Miami, FL)	1.18	1.18	NA	0.81	0.86	1,626	1,626	NA	1,812	1,875	6.2	6.2	NA	5.8	5.0
Benfield Transitway (Portland, OR)	1.17	0.59	NA	0.83	1.02	563	560	NA	575	582	9.2	7.5	NA	8.4	5.4
Light Rail (Pittsburgh, PA)	0.62	0.62	NA	0.85	0.94	1,481	1,481	NA	1,371	1,348	6.8	6.8	NA	5.6	4.6
People Mover (Detroit, MI)	1.21	1.21	NA	0.86	0.86	2,338	2,338	NA	2,180	2,157	13.5	13.5	NA	8.4	7.4
Light Rail (Sacramento, CA)	1.36	1.36	NA	0.84	0.92	809	809	NA	964	1,015	7.1	7.1	NA	5.7	5.1
El Cajon Extension (San Diego, CA)	1.05	1.05	0.80	0.92	1.51	2,180	2,180	2,244	2,434	2,825	7.3	7.3	6.8	5.1	5.0
Downtown Transit (Seattle, WA)	1.12	NA	0.88	1.14	1.10	1,389	NA	1,412	1,517	1,541	8.4	NA	8.2	5.1	6.3
Guadalupe (San Jose, CA)	1.36	1.36	1.05	0.98	1.11	1,315	1,315	1,376	1,513	1,532	7.1	7.1	7.7	7.8	9.3
Orange Line (Chicago, IL)	1.24	1.24	0.82	1.04	1.48	5,224	5,224	5,164	5,235	5,377	11.3	11.3	8.4	7.7	4.3
SW Transitway (Houston, TX)	1.06	1.06	0.87	1.08	1.24	2,614	2,614	2,696	3,034	3,558	7.1	7.1	8.5	7.1	6.4
Metrolink (St Louis, MO)	1.02	1.02	0.82	0.99	1.01	429	429	412	376	369	6.9	6.9	5.9	4.8	4.9
Metromover Extensions (Miami, FL)	0.86	0.86	0.92	1.00	1.33	1,844	1,844	1,905	2,045	2,259	5.2	5.2	5.6	6.5	3.8
I-25 Transitway (Denver, CO)	0.98	NA	0.98	1.11	1.49	473	NA	473	513	556	5.6	NA	5.6	4.3	2.7
Metro Extension (Baltimore, MD)	1.08	1.08	1.04	1.26	1.40	768	768	748	702	645	5.4	5.4	4.3	5.1	3.9
Colma station (San Francisco, CA)	0.85	0.85	1.10	1.21	1.51	730	730	740	754	777	5.3	5.3	9.4	7.3	5.0
South Oak Cliff (Dallas, TX)	1.10	1.10	1.08	1.15	1.24	1,864	1,864	1,970	2,119	2,280	6.3	6.3	7.1	5.3	6.4
LRT Extensions (Baltimore, MD)	1.19	1.19	1.22	1.29	1.40	736	736	713	677	645	6.0	6.0	5.2	4.9	3.9
Hillsboro Extension (Portland, OR)	1.21	1.21	1.21	1.13	1.38	617	617	622	652	676	7.2	7.2	5.4	5.8	7.3
N/S LRT (Salt Lake City, UT)	1.09	1.09	1.11	1.22	1.33	730	730	836	891	925	4.4	4.4	3.5	3.7	5.9
Tasman West LRT (San Jose, CA)	0.98	0.98	1.21	1.26	1.34	1,513	1,513	1,609	1,671	1,673	7.8	7.8	7.3	5.2	6.7
Skyway Express (Jacksonville, FL)	1.27	1.27	1.09	1.33	1.25	591	591	622	780	780	8.3	8.3	6.1	3.8	3.8
Red Line (Los Angeles, CA)	NA	0.92	NA	1.51	1.47	NA	8,794	NA	9,543	9,635	NA	5.1	NA	5.0	5.3
Red Line MOS 1 (Los Angeles, CA)	NA	NA	1.05	1.10	NA	NA	NA	8,183	9,100	NA	NA	NA	7.3	9.4	NA
Red Line MOS 2 (Los Angeles, CA)	NA	NA	1.03	1.26	NA	NA	NA	8,878	9,437	NA	NA	NA	5.5	5.2	NA
Red Line MOS 3 (Los Angeles, CA)	NA	NA	1.23	1.51	NA	NA	NA	9,207	9,543	NA	NA	NA	6.3	5.0	NA
North Line Extension (Atlanta, GA)	0.99	0.99	0.91	1.25	1.30	651	651	717	817	882	5.3	5.3	5.2	3.7	5.1

Table B-7: External control variables for study sample (2/3)

	State average gasoline price at time of…					County population (thousands) at time of…					State unemployment rate (June, seasonally adjusted) at time of…				
Project name	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership
Airport Busway (Pittsburgh, PA)	1.12	1.12	1.11	1.46	1.29	1,341	1,341	1,332	1,280	1,265	7.7	7.7	6.3	4.1	5.6
Southwest Light Rail (Denver, CO)	1.19	1.19	1.17	1.49	1.37	513	513	549	556	561	4.3	4.3	3.1	2.7	5.6
University Ext (Salt Lake City, UT)	NA	NA	1.48	1.41	2.18	NA	NA	901	914	960	NA	NA	3.3	4.3	4.0
St Clair Extension (St Louis, MO)	0.99	0.99	1.12	1.31	1.24	376	376	362	350	351	4.8	4.8	4.8	4.3	5.5
North Central Ext (Dallas, TX)	1.08	1.08	1.07	1.24	2.62	1,999	1,999	2,073	2,280	2,382	6.6	6.6	4.7	6.4	4.2
BART to SFO (San Francisco, CA)	1.10	1.10	1.23	1.66	2.80	740	740	763	775	799	9.4	9.4	6.3	6.9	5.3
Med Center Ext (Salt Lake City, UT)	NA	NA	1.33	1.55	2.18	NA	NA	925	936	960	NA	NA	5.9	5.7	4.0
South LRT (Sacramento, CA)	1.11	1.11	1.23	1.66	2.80	1,141	1,141	1,170	1,325	1,374	8.0	8.0	6.3	6.9	5.3
Piers Transitway (Boston, MA)	1.15	1.15	1.15	1.81	2.71	1,347	1,347	1,354	1,500	1,512	6.9	6.9	6.0	5.2	4.6
Largo Extension (Washington, DC)	1.36	1.36	1.45	1.80	2.72	572	572	572	568	574	8.7	8.7	5.4	7.9	5.5
Interstate MAX (Portland, OR)	1.35	1.35	1.58	1.90	2.86	640	640	662	672	697	5.7	5.7	5.1	7.3	5.1
Hiawatha (Minneapolis, MN)	1.17	1.17	1.77	1.77	2.71	1,110	1,110	1,123	1,123	1,134	2.9	2.9	4.8	4.8	4.6
Med Center Ext (Memphis, TN)	0.99	0.99	1.37	1.72	2.62	886	886	898	909	920	4.3	4.3	3.9	5.2	4.4
Tren Urbano (San Juan, PR)	NA	NA	NA	NA	NA	437	437	436	420	410	17.2	17.2	14.4	11.2	11.3
Reconstruction II (Pittsburgh, PA)	1.11	1.11	1.35	2.19	2.73	1,332	1,332	1,273	1,232	1,220	6.3	6.3	4.8	4.9	4.4
Mission Valley East (San Diego, CA)	1.08	1.08	1.51	2.28	2.80	2,737	2,737	2,825	2,942	2,976	5.9	5.9	5.0	5.3	5.3
Douglas Reconstr. (Chicago, IL)	1.12	1.12	1.44	2.13	2.70	5,365	5,365	5,377	5,269	5,236	4.5	4.5	5.2	5.7	4.9
Elizabeth Link (Newark, NJ)	1.07	1.07	1.41	2.41	3.07	790	790	792	772	767	7.6	7.6	3.6	4.9	5.1
Hudson Bergen I (Jersey City, NJ)	NA	1.09	1.16	1.24	3.07	NA	564	585	608	592	NA	8.7	6.1	5.8	5.1
Hudson Bergen II (Jersey City, NJ)	NA	1.09	1.41	2.41	3.07	NA	564	609	592	592	NA	8.7	3.6	4.9	5.1
North Central (Chicago, IL)	1.12	1.12	1.44	2.43	2.43	5,346	5,346	5,377	5,237	5,237	4.3	4.3	5.2	4.5	4.5
Southwest (Chicago, IL)	1.12	1.12	1.44	2.43	2.43	5,346	5,346	5,377	5,237	5,237	4.3	4.3	5.2	4.5	4.5
UP West Ext (Chicago, IL)	1.12	1.12	1.44	2.43	2.43	5,346	5,346	5,377	5,237	5,237	4.3	4.3	5.2	4.5	4.5
Southeast Corridor (Denver, CO)	1.07	1.07	1.49	2.49	2.76	541	541	556	569	579	3.7	3.7	2.7	4.3	3.6
Central Double Track (Baltimore)	1.19	1.19	1.40	2.52	3.14	657	657	645	641	638	3.6	3.6	3.9	4.0	4.1
Ft Lauderdale Dbl-Track (Miami)	1.02	1.02	1.33	2.59	2.59	2,221	2,221	2,259	2,454	2,454	3.9	3.9	3.8	4.0	4.0
South Corridor (Charlotte, NC)	1.06	0.98	2.19	2.69	2.28	682	660	800	866	914	3.2	3.2	5.3	4.7	10.7
Sprinter light rail (Oceanside, CA)	1.11	1.11	1.34	3.23	2.98	2,624	2,624	2,901	3,019	3,104	8.0	8.0	6.7	7.2	12.1
Euclid Corridor (Cleveland, OH)	1.18	1.18	1.77	3.12	2.72	1,411	1,411	1,341	1,283	1,278	4.5	4.5	6.2	6.4	10.1
East Valley (Phoenix, AZ)	1.07	1.07	2.25	3.11	3.33	2,909	2,909	3,647	3,958	3,869	4.2	4.2	4.6	5.9	9.7
Frontrunner N (Salt Lake City, UT)	1.55	1.55	2.46	3.16	2.32	936	936	985	1,019	1,035	5.7	5.7	2.9	3.3	7.6

Table B-7: External control variables for study sample (3/3)

Project name	State average gasoline price at time of…					County population at time of…					State unemployment at time of…				
	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership
Gold Line Ext. (Los Angeles, CA)	1.51	1.51	1.96	2.47	3.67	9,543	9,543	9,808	9,848	9,885	5.0	5.0	6.2	11.3	11.7
Central, Airport Links (Seattle)	1.55	1.55	1.47	2.55	3.67	1,739	1,739	1,757	1,916	1,971	5.1	5.1	6.1	9.3	9.3
Westside Express (Portland, OR)	1.50	1.50	2.59	2.89	3.74	670	670	683	727	726	6.2	6.2	5.3	11.8	11.8
Green Line, (Portland, OR)	1.90	1.90	2.86	2.47	3.56	672	672	697	727	748	7.3	7.3	5.1	11.8	9.5
Northstar (Minneapolis, MN)	2.12	2.12	2.71	2.30	3.53	1,122	1,122	1,134	1,156	1,169	3.9	3.9	4.6	8.1	6.6
NW-SE Line (Dallas, TX)	1.29	1.29	2.40	2.61	3.39	2,268	2,268	2,353	2,373	2,454	4.8	4.8	5.0	8.0	6.7
The Tide (Norfolk, VA)	1.26	1.26	2.60	3.36	3.36	240	240	236	244	246	4.2	4.2	3.0	6.5	5.6
Mountainlink (Flagstaff, AZ)	2.25	2.25	3.33	3.33	3.37	130	130	134	134	137	10.3	10.3	9.7	9.7	7.5

Table B-8: Sources of external control variables for study sample (1/3)

Project name	State average gasoline price at time of…					County population at time of…					June central city unemployment at time of…				
	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership
Heavy Rail (Baltimore, MD)	(41)	(41)	NA	(41)	(41)	(42)	(42)	NA	(42)	(42)	(44)	(44)	NA	(44)	(44)
Heavy Rail (Washington, DC)	NA	NA	NA	(41)	(41)	(42)	(42)	NA	(42)	(42)	(44)	(44)	NA	(44)	(44)
Heavy Rail (Miami, FL)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
Heavy Rail (Atlanta, GA)	(41)	NA	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
Light Rail (Buffalo, NY)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
People Mover (Miami, FL)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
Benfield Transitway (Portland, OR)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
Light Rail (Pittsburgh, PA)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
People Mover (Detroit, MI)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
Light Rail (Sacramento, CA)	(41)	(41)	NA	(41)	(41)	(43)	(43)	NA	(43)	(43)	(44)	(44)	NA	(44)	(44)
El Cajon Extension (San Diego, CA)	(41)	(41)	(41)	(41)	(41)	(43)	(43)	(43)	(43)	(42)	(44)	(44)	(44)	(44)	(44)
Downtown Transit (Seattle, WA)	(41)	NA	(41)	(41)	(41)	(43)	NA	(43)	(42)	(42)	(44)	NA	(44)	(44)	(44)
Guadalupe (San Jose, CA)	(41)	(41)	(41)	(41)	(41)	(43)	(43)	(43)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Orange Line (Chicago, IL)	(41)	(41)	(41)	(41)	(41)	(43)	(43)	(43)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
SW Transitway (Houston, TX)	(41)	(41)	(41)	(41)	(41)	(43)	(43)	(43)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Metrolink (St Louis, MO)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Metromover Extensions (Miami, FL)	(41)	(41)	(41)	(41)	(41)	(43)	(43)	(43)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
I-25 Transitway (Denver, CO)	(41)	NA	(41)	(41)	(41)	(42)	NA	(42)	(42)	(42)	(44)	NA	(44)	(44)	(44)
Metro Extension (Baltimore, MD)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Colma station (San Francisco, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
South Oak Cliff (Dallas, TX)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
LRT Extensions (Baltimore, MD)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Hillsboro Extension (Portland, OR)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
N/S LRT (Salt Lake City, UT)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Tasman West LRT (San Jose, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Skyway Express (Jacksonville, FL)	(41)	(41)	(41)	(41)	(41)	(43)	(43)	(43)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Red Line (Los Angeles, CA)	NA	(41)	NA	(41)	(41)	NA	(43)	NA	(42)	(42)	NA	(44)	NA	(44)	(44)
Red Line MOS 1 (Los Angeles, CA)	NA	NA	(41)	(41)	NA	NA	NA	(43)	(42)	NA	NA	NA	(44)	(44)	NA
Red Line MOS 2 (Los Angeles, CA)	NA	NA	(41)	(41)	NA	NA	NA	(43)	(42)	NA	NA	NA	(44)	(44)	NA
Red Line MOS 3 (Los Angeles, CA)	NA	NA	(41)	(41)	NA	NA	NA	(42)	(42)	NA	NA	NA	(44)	(44)	NA
North Line Extension (Atlanta, GA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)

Table B-8: Sources of external control variables for study sample (2/3)

	State average gasoline price at time of…					County population at time of…					June central city unemployment at time of…				
Project name	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership
Airport Busway (Pittsburgh, PA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Southwest Light Rail (Denver, CO)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
University Ext (Salt Lake City, UT)	NA	NA	(41)	(41)	(41)	NA	NA	(42)	(42)	(42)	NA	NA	(44)	(44)	(44)
St Clair Extension (St Louis, MO)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
North Central Ext (Dallas, TX)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
BART to SFO (San Francisco, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Med Center Ext (Salt Lake City, UT)	NA	NA	(41)	(41)	(41)	NA	NA	(42)	(42)	(42)	NA	NA	(44)	(44)	(44)
South LRT (Sacramento, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Piers Transitway (Boston, MA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Largo Extension (Washington, DC)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Interstate MAX (Portland, OR)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Hiawatha (Minneapolis, MN)	NA	NA	(41)	(41)	(41)	NA	NA	(42)	(42)	(42)	NA	NA	(44)	(44)	(44)
Med Center Ext (Memphis, TN)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Tren Urbano (San Juan, PR)	NA	NA	NA	NA	NA	(43)	(43)	(43)	(42)	(42)	(45)	(45)	(45)	(45)	(45)
Reconstruction II (Pittsburgh, PA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Mission Valley East (San Diego, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Douglas Reconstr. (Chicago, IL)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Elizabeth Link (Newark, NJ)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Hudson Bergen I (Jersey City, NJ)	NA	(41)	(41)	(41)	(41)	NA	(42)	(42)	(42)	(42)	NA	(44)	(44)	(44)	(44)
Hudson Bergen II (Jersey City, NJ)	NA	(41)	(41)	(41)	(41)	NA	(42)	(42)	(42)	(42)	NA	(44)	(44)	(44)	(44)
North Central (Chicago, IL)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Southwest (Chicago, IL)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
UP West Ext (Chicago, IL)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Southeast Corridor (Denver, CO)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Central Double Track (Baltimore)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Ft Lauderdale Dbl-Track (Miami)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
South Corridor (Charlotte, NC)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Sprinter light rail (Oceanside, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Euclid Corridor (Cleveland, OH)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
East Valley (Phoenix, AZ)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Frontrunner N (Salt Lake City, UT)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)

Table B-8: Sources of external control variables for study sample (3/3)

Project name	State average gasoline price at time of…					County population at time of…					June central city unemployment at time of…				
	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership	Initial cost forecast	Rider-ship forecast	Final cost forecast	Project opening	Observed ridership
Gold Line Ext. (Los Angeles, CA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Central, Airport Links (Seattle)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Westside Express (Portland, OR)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Green Line, (Portland, OR)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Northstar (Minneapolis, MN)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
NW-SE Line (Dallas, TX)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
The Tide (Norfolk, VA)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)
Mountainlink (Flagstaff, AZ)	(41)	(41)	(41)	(41)	(41)	(42)	(42)	(42)	(42)	(42)	(44)	(44)	(44)	(44)	(44)

Data sources:
(1) Pickrell (1989)
(2) Kozel (2002)
(3) Schrag (2008)
(4) Miami-Dade County (2015)
(5) Metropolitan Atlanta Rapid Transit Authority (2005)
(6) Niagara Frontier Transportation Authority (2010)
(7) Spielberg et al. (2003)
(8) Federman (1982)
(9) Fisher (1980)
(10) Detroit Transportation Corporation (2007)
(11) Bee Metro Staff (1984)
(12) Lewis-Workman et al. (2008)
(13) Federal Transit Administration (2011b)
(14) Federal Transit Administration (2012b)
(15) Federal Transit Administration (2013b)
(16) Federal Transit Administration (2015b)
(17) Federal Transit Administration (2016)
(18) Estimate in nominal dollars based on values in Pickrell (1989)
(19) Estimated based on Lewis-Workman et al.(2008).
(20) Federal contribution from Urban Mass Transportation Administration (1990)
(21) San Diego Association of Governments (1985)
(22) Federal Transit Administration (1994b), adjusted for double-tracking
(23) Federal contribution from Federal Transit Administration (1994a)
(24) Federal contribution from Urban Mass Transportation Administration (1991)
(25) Federal contribution from Federal Transit Administration (1993)
(26) Federal contribution from Federal Transit Administration (1996a)
(27) Federal contribution from Federal Transit Administration (2001)
(28) Federal contribution from Federal Transit Administration (1997)
(29) Federal contribution from Federal Transit Administration (1999)
(30) Denver Regional Transportation District (2016)
(31) Federal contribution from Federal Transit Administration (2006)
(32) Federal contribution from Federal Transit Administration (2002)
(33) Federal contribution from Federal Transit Administration (2000b)
(34) Federal contribution from Federal Transit Administration (2008)
(35) Washington Metropolitan Area Transit Authority (2016)
(36) Memphis Business Journal Staff (2009)
(37) Federal contribution from Federal Transit Administration (2003)
(38) Chicago Metra (2015)
(39) Los Angeles County Metropolitan Transportation Authority (2016)
(40) Dallas Area Rapid Transit (2016)
(41) Price per million British thermal units (Btu) from U.S. Energy Information Administration (2016), converted to price per gallon based on 120,405 Btu/gallon (U.S. Energy Information Administration 2016)
(42) Google Public Data Explorer (2016)
(43) U.S. Census Bureau (2000), interpolated between decennial years
(44) Federal Reserve Bank of St Louis (2016)
(45) Bureau of Labor Statistics (2016)

Bibliography

109th Congress. 2005. Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users.

112th Congress. 2012. Moving Ahead for Progress in the 21st Century.

114th Congress. 2015. Fixing America's Surface Transportation Act.

Ackerman, Andy. 1995. "The Beard." Seinfeld, Season 6. NBC.

Alonso, William. 1968. "Predicting Best with Imperfect Data." Journal of the American Institute of Planners 34 (4): 248–55. doi:10.1080/01944366808977813.

Altshuler, Alan A., and David E. Luberoff. 2003. Mega-Projects: The Changing Politics of Urban Public Investment. Washington, D.C.; Cambridge, Mass.: Brookings Institution Press and the Lincoln Institute of Land Policy.

Armstrong, J. Scott. 1978. Long-Range Forecasting: From Crystal Ball to Computer. 1st ed. New York: John Wiley & Sons, Inc.

Armstrong, J. Scott, and Fred Collopy. 1992. "Error Measures for Generalizing about Forecasting Methods: Empirical Comparisons." International Journal of Forecasting 8 (1): 69–80. doi:10.1016/0169-2070(92)90008-W.

Ascher, William. 1979. Forecasting: An Appraisal for Policy-Makers and Planners. Baltimore: Johns Hopkins University Press.

Ashworth, A., and R.M. Skitmore. 2005. "Accuracy in Estimating." In Cost Modelling, edited by Martin Skitmore and Vernon Marston. Routledge.

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

Bayes, Thomas, and Richard Price. 1763. "An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a

Letter to John Canton, A. M. F. R. S.” Philosophical Transactions (1683-1775) 53: 370–418.

Bee Metro Staff. 1984. “New K Street Mall Trees Planned along Light Rail Foliage Lost to Trolleys Wil Be Replaced.” Sacramento Bee, April 5, sec. Metro.

Boyle, Daniel. 2006. “Fixed-Route Transit Ridership Forecasting and Service Planning Methods.” TCRP Synthesis 66. Transit Cooperative Research Program. http://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_syn_66.pdf.

Boyle, Melissa A., and Victor A. Matheson. 2009. “Determinants of the Distribution of Congressional Earmarks across States.” Economics Letters 104 (2): 63–65. doi:10.1016/j.econlet.2009.04.001.

Bureau of Labor Statistics. 2016. “Local Area Unemployment Statistics.” http://www.bls.gov/lau/.

Button, Kenneth. 2009. “Transit Forecasting Accuracy: Ridership Forecasts and Capital Cost Estimates.” Washington DC: George Mason University. https://www.researchgate.net/profile/Kenneth_Button/publication/268056254_Transit_Forecasting_Accuracy_Ridership_Forecasts_and_Capital_Cost_Estimates/links/54c668a80cf256ed5a9e63db.pdf.

Chicago Metra. 2015. “Operations and Ridership Data.” About Metra. https://metrarail.com/about-metra/reports-documents/operations-ridership-data.

Dallas Area Rapid Transit. 2016. Dallas Area Rapid Transit Reference Book. 7.0 ed. Dallas: Dallas Area Rapid Transits. https://www.dart.org/about/dartreferencebookmar16.pdf.

Dantata, N. A., A. Touran, and D. C. Schneck. 2006. “Trends in U.S. Rail Transit Project Cost Overrun.” In 2006 Annual Meeting CD-ROM. Washington DC.

Davis, Jeff. 2015. “Reagan Devolution: The Real Story of the 1982 Gas Tax Increase.” Eno Center for Transportation.

De Bruijn, Hans, and Martijn Leijten. 2007. “Megaprojects and Contested Information.” Transportation Planning and Technology 30 (1): 49–69. doi:10.1080/03081060701208050.

Demery Jr., Leroy, and Michael D. Setty. 2007. “Flyvbjerg, Skamiris Holm and Buhl: Deja va All Over Again?” Special Report 5. publictransit.us. http://www.publictransit.us/ptlibrary/specialreports/sr5.Flyvbjerg_Deja_Vue.htm.

Denver Regional Transportation District. 2016. “Central Corridor Light Rail Line.” RTD Facts and Figures. January. http://www.rtd-denver.com/FF-CentralCorridorLRT.shtml.

Detroit Transportation Corporation. 2007. “About the Detroit People Mover.” Detroit Transportation Corporation. http://www.thepeoplemover.com/resource/attach/26/GeneralInformationBrochureFinal.pdf.

Dorn, Jennifer L. Dear Colleague letter. 2005. “Changes to the New Starts Rating Process,” April 29. http://www.fta.dot.gov/newsroom_300.html.

Doyle, Richard B. 2011. “The Rise and (Relative) Fall of Earmarks: Congress and Reform, 2006–2010.” Public Budgeting & Finance 31 (1): 1–22. doi:10.1111/j.1540-5850.2011.00971.x.

Duff, Daniel, Edward J. Gill Jr., G. Kent Woodman, and Thompson Coburn. 2010. “Legal Handbook for the New Starts Process.” TCRP Legal Research Digest, no. 30 (February). http://trid.trb.org/view.aspx?id=914287.

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

Elkind, Ethan N. 2014. Railtown: The Fight for the Los Angeles Metro Rail and the Future of the City. 1 edition. University of California Press.

Emerson, Don. 2012. “What’s New in New Starts: The Map-21 Reforms.” EFR 6 (2). http://efr.wsp-pb.com/publications/default.aspx?id=14.

Fair, Ray C. 1971. A Short-Run Forecasting Model of the United States. Lexington, Massachusetts: Lexington Books, D.C. Heath and Company.

Federal Highway Administration. 1995. “A Guide to Metropolitan Transportation Planning Under ISTEA: How the Pieces Fit Together.” Washignton DC. https://trid.trb.org/view.aspx?id=423085.

Federal Reserve Bank of St Louis. 2016. “Federal Reserve Economic Data.” https://fred.stlouisfed.org/.

Federal Transit Administration. 1993. “Report on Funding Levels and Allocations of Funds: Report of the Secretary of Transportation to the United States Congress Pursuant to Section 3 (j) of the Federal Transit Act, as Amended.” Washington, D.C.: United States Department of Transportation.

———. 1994a. “Report on Funding Levels and Allocations of Funds: Report of the Secretary of Transportation to the United States Congress pursuant to Section 3(j) of the Federal Transit Act, as Ammended.” Washington DC: United States Department of Transportation. http://hdl.handle.net/2027/uc1.l0072776867.

———. 1994b. Data Tables: For the 1993 National Transit Database Section 15 Report Year. Federal Transit Administration/US Department of Transportation. http://archive.org/details/datatablesfor19900gord_3.

———. 1996a. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(m)(3) (Formerly Section 3(j) of the Federal Transit Act).” Washington DC.

https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY97_Annual_Report_on_Funding_Recommendations.pdf.

———. 1996b. “Section 5309 (Section 3(j)) FTA New Starts Criteria.” Federal Register 61 (245): 67093–106.

———. 1997. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(m)(3) (Formerly Section 3(j) of the Federal Transit Act).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY98_Annual_Report_on_Funding_Recommendations.pdf.

———. 1998. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(m)(3) (Formerly Section 3(j) of the Federal Transit Act).” Washington DC.

———. 1999. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(o)(1).” FTA-TBP10-99-01. Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY00_Annual_Report_on_Funding_Recommendations.pdf.

———. 2000a. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1).” Washington DC.

———. 2000b. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1).” FTA-TBP10-00-01. Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY01_Annual_Report_on_Funding_Recommendations.pdf.

———. 2000c. "Major Capital Investment Projects: Final Rule." Federal Register 65 (236): 76864–84.

———. 2001. "Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1)." FTA-TBP10-01-01. Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY02_Annual_Report_on_Funding_Recommendations.pdf.

———. 2002. "Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1)." FTA-TBP10-2002-01. Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY03_Annual_Report_on_Funding_Recommendations.pdf.

———. 2003. "Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1)." FTA-TBP10-2003-02. Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY04_Annual_Report_on_Funding_Recommendations.pdf.

———. 2004. "Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1)." FTA-TBP10-2004-1. Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY05_Annual_Report_on_Funding_Recommendations.pdf.

———. 2005. "Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309 (o)(1)." FTA-TBP10-2005-1. Washington DC.

https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY06_Annual_Report_on_Funding_Recommendations.pdf.

———. 2006. “Annual Report on New Starts Proposed Allocations of Funds Fiscal Year 2007: Report of the Secretary of Transportation to the United States Congress Pursuant to 40 U.S.C. 5309(k)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY07_NEW_STARTS_REPORT_COMPLETE.pdf.

———. 2007. “Annual Report on New Starts Proposed Allocations of Funds Fiscal Year 2008.” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY2008_Entire_NS_Report.pdf.

———. 2008. “Annual Report on Funding Recommendations: Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(k)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY09_Annual_Report_on_Funding_Recommendations.pdf.

———. 2009. “Annual Report on Funding Recommendations Fiscal Year 2010: New Starts, Small Starts, and Paul S. Sarbanes.” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/FY10_Annual_Report_on_Funding_Recommendations.pdf.

———. 2010. “Annual Report on Funding Recommendations Fiscal Year 2011: New Starts, Small Starts, and Paul S. Sarbanes Transit in Parks Program. Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(k)(1).” Washington DC.

https://www.transit.dot.gov/sites/fta.dot.gov/files/FY11_Annual_Report_on_Funding_Recommendations_REVISED.pdf.

———. 2011a. “Annual Report on Funding Recommendations Fiscal Year 2012: Capital Investment and Paul S. Sarbanes Transit in Parks Programs. Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(k)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/FY12_Annual_Report_on_Funding_Recommendations.pdf.

———. 2011b. “Before-and-After Studies of New Starts Projects: Report to Congress.”

———. 2012a. “Annual Report on Funding Recommendations Fiscal Year 2013: Capital Investment and Paul S. Sarbanes Transit in Parks Programs. Report of the Secretary of Transportation to the United States Congress Pursuant to 49 U.S.C. 5309(k)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/FY13_Annual_Report_on_Funding_Recommendations.pdf.

———. 2012b. “Before-and-After Studies of New Starts Projects: Report to Congress.”

———. 2013a. “Annual Report on Funding Recommendations Fiscal Year 2014: Capital Investment Grant Program. Report of the Secretary of Transportation to the United States Congress Pursuant to 49 USC 5309(o)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/FY14_Annual_Report_on_Funding_Recommendations.pdf.

———. 2013b. “Before-and-After Studies of New Starts Projects: Report to Congress.”

———. 2013c. “Major Capital Investment Projects.” Federal Register 78 (6): 1992–2037.

———. 2013d. “An Overview of STOPS.” https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/STOPS.overview-web-final.pdf.

———. 2014. “Annual Report on Funding Recommendations Fiscal Year 2015: Capital Investment Grant Program. Report of the Secretary of Transportation to the United States Congress Pursuant to 49 USC 5309(o)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/FY15_Annual_Report_on_Funding_Recommendations.pdf.

———. 2015a. “Annual Report on Funding Recommendations Fiscal Year 2016: Capital Investment Grant Program. Report of the Secretary of Transportation to the United States Congress Pursuant to 49 USC 5309(o)(1).” Washington DC. https://www.transit.dot.gov/sites/fta.dot.gov/files/FY16_Annual_Report_on_Funding_Recommendations.pdf.

———. 2015b. “Before-and-After Studies of New Starts Projects: Report to Congress.”

———. 2016. “Before-and-After Studies of New Starts Projects: Report to Congress.”

Federman, Stan. 1982. “At Ground-Breaking Festivities Herald Transitway.” The Oregonian, March 27, sec. Metro/Northwest.

Ferry, Douglas J., Peter S. Brandon, and Richard Kirkham. 2014. Ferry and Brandon’s Cost Planning of Buildings. 9 edition. Chichester, West Sussex, United Kingdom: Wiley-Blackwell.

Fisher, Ken. 1980. “Bill Stalled, Extra Funds for PAT Seem Doomed.” Pittsburgh Post-Gazette, December 10, sec. City/Area.

Flanagan, Roger, and George Norman. 1983. “The Accuracy and Monitoring of Quantity Surveyors’ Price Forecasting for Building Work.” Construction Management and Economics 1 (2): 157–80. doi:10.1080/01446198300000012.

Fleming, Victor. 1939. Wizard of Oz. Fantasy. Metro-Goldwyn-Mayer.

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. doi:10.1016/j.tra.2005.02.003.

———. 2006. “From Nobel Prize to Project Management: Getting Risks Right.” Project Management Journal 37 (3). http://arxiv.org/abs/1302.3642.

———. 2008. “Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice.” European Planning Studies 16 (1): 3–21. doi:10.1080/09654310701747936.

———. 2016. “The Fallacy of Beneficial Ignorance: A Test of Hirschman’s Hiding Hand.” World Development 84 (April). http://papers.ssrn.com/abstract=2767128.

Flyvbjerg, Bent, Nils Bruzelius, and Werner Rothengatter. 2003. Megaprojects and Risk: An Anatomy of Ambition. Cambridge University Press.

Flyvbjerg, Bent, Mette Skamris Holm, and Soren Buhl. 2002. “Underestimating Costs in Public Works Projects: Error or Lie?” Journal of the American Planning Association 68 (3): 279–95. doi:10.1080/01944360208976273.

Flyvbjerg, Bent, Mette K. Skamris Holm, and Søren L. Buhl. 2003. “How Common and How Large Are Cost Overruns in Transport Infrastructure Projects?” Transport Reviews 23 (1): 71–88. doi:10.1080/01441640309904.

———. 2005. “How (In)accurate Are Demand Forecasts in Public Works Projects?: The Case of Transportation.” Journal of the American Planning Association 71 (2): 131–46. doi:10.1080/01944360508976688.

General Accounting Office. 2000. “Project Management Oversight Benefits and Future Funding Requirements.” GAO/RCED-00-221. Report to Congressional Committees: Mass Transit. Washington DC. http://www.gao.gov/assets/240/230604.pdf.

———. 2003. “FTA Needs to Provide Clear Information and Additional Guidance on the New Starts Ratings Process.” GAO-03-701. Report to Congressional Committees: Mass Transit. Washington DC. http://www.gao.gov/assets/240/238445.pdf.

———. 2005. “Opportunities Exist to Improve the Communication and Transparency of Changes Made to the New Starts Program.” GAO-05-674. Report to Congressional Committees: Public Transportation. Washington DC. http://www.gao.gov/assets/250/246862.pdf.

Gneiting, Tilmann. 2011. “Making and Evaluating Point Forecasts.” Journal of the American Statistical Association 106 (494): 746–62. doi:10.1198/jasa.2011.r10138.

Gomez-Ibanez, Jose A. 1985a. “The Federal Role in Urban Transportation.” In American Domestic Priorities: An Economic Appraisal, edited by John M. Quigley and Daniel L. Rubinfeld. Berkeley: University of California Press.

———. 1985b. “A Dark Side to Light Rail? The Experience of Three New Transit Systems.” Journal of the American Planning Association 51 (3): 337–51. doi:10.1080/01944368508976421.

Google Public Data Explorer. 2016. “Population in the U.S.” October 11. https://www.google.com/publicdata/explore?ds=kf7tgg1uo9ude_.

Healy, M. J. R., and H. Goldstein. 1978. “Regression to the Mean.” Annals of Human Biology 5 (3): 277–80. doi:10.1080/03014467800002891.

Hickman, Bert G. 1972. “Introduction and Summary.” In Econometric Models of Cyclical Behavior, edited by Bert G. Hickman. Vol. 2. Studies in Income and Wealth 36. New York: Columbia University Press.

Hinze, Jimmie, Gregory Selstead, and Joe P. Mahoney. 1992. “Cost Overruns on State of Washington Construction Contracts.” Transportation Research Record, no. 1351.

Hirschman, Albert O. 1967. “The Principle of the Hiding Hand.” The Public Interest 0 (6): 10–23.

Hofstadter, Douglas. 1985. “The Prisoner’s Dilemma, Computer Tournaments and the Evolution of Cooperation.” In Metamagical Themas: Questing For The Essence Of Mind And Pattern, Book Club, 715–38. Basic Books.

Hood, Clifton. 2004. 722 Miles: The Building of the Subways and How They Transformed New York. Reprint edition. Baltimore, Md: Johns Hopkins University Press.

Hyndman, Rob J., and Anne B. Koehler. 2006. “Another Look at Measures of Forecast Accuracy.” International Journal of Forecasting 22 (4): 679–88. doi:10.1016/j.ijforecast.2006.03.001.

Jahren, C., and A. Ashe. 1990. “Predictors of Cost-Overrun Rates.” Journal of Construction Engineering and Management 116 (3): 548–52. doi:10.1061/(ASCE)0733-9364(1990)116:3(548).

Kahneman, Daniel. 2013. Thinking, Fast and Slow. Reprint edition. New York: Farrar, Straus and Giroux.

Kahneman, Daniel, and A Tversky. 1979. “Intuitive Prediction: Biases and Corrective Procedures.” Management Science 12: 313?327–313?327.

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. doi:10.1080/01944369008975758.

Kennedy, John F. 1963. Public Papers of the Presidents of the United States: John F. Kennedy, 1962. U.S. Government Printing Office.

Keyfitz, Nathan. 1981. “The Limits of Population Forecasting.” Population and Development Review 7 (4): 579–93. doi:10.2307/1972799.

Kitchenham, B. A., L. M. Pickard, S. G. MacDonell, and M. J. Shepperd. 2001. “What Accuracy Statistics Really Measure [software Estimation].” IEE Proceedings - Software 148 (3): 81–85. doi:10.1049/ip-sen:20010506.

Kozel, Scott M. 2002. “Baltimore Metro Subway.” Roads to the Future. October 13. http://www.roadstothefuture.com/Baltimore_Metro.html.

LaPlace, Pierre Simon de. 1902. A Philisophical Essay on Probabilities. Translated by Frederick Wilson Truscott and Frederick Lincoln Emory. First. New York: John Wiley & Sons, Ltd.

Lave, Charles A. 1985. “Speeding, Coordination, and the 55 MPH Limit.” The American Economic Review 75 (5): 1159–64.

Lewis-Workman, Steven, Bryon White, Stephanie McVey, and Frank Spielberg. 2008. “The Predicted and Actual Impacts of New Starts Projects.” Washington DC: US Department of Transportation Federal Transit Administration.

Liu, Li, and Kai Zhu. 2007. “Improving Cost Estimates of Construction Projects Using Phased Cost Factors.” Journal of Construction Engineering and Management 133 (1): 91–95. doi:10.1061/(ASCE)0733-9364(2007)133:1(91).

Lo, Bruce, and Xiangzhu Gao. 1997. “Assessing Software Cost Estimation Models: Criteria for Accuracy, Consistency and Regression.” Australasian Journal of Information Systems 5 (1). http://journal.acs.org.au/index.php/ajis/article/view/348.

Lokan, C. 2005. “What Should You Optimize When Building an Estimation Model?” In 11th IEEE International Software Metrics Symposium (METRICS’05), 10 pp. – 34. doi:10.1109/METRICS.2005.55.

Los Angeles County Metropolitan Transportation Authority. 2016. “Facts at a Glance.” News & Media Now. https://www.metro.net/news/facts-glance/.

Love, Peter E. D., Chun-pong Sing, Brad Carey, and Jeong Tai Kim. 2015. “Estimating Construction Contingency: Accommodating the Potential for Cost Overruns in Road Construction Projects.” Journal of Infrastructure Systems 21 (2): 04014035. doi:10.1061/(ASCE)IS.1943-555X.0000221.

Love, Peter E. D., Xiangyu Wang, Chun-pong Sing, and Robert L. K. Tiong. 2013. “Determining the Probability of Project Cost Overruns.” Journal of Construction Engineering and Management 139 (3): 321–30. doi:10.1061/(ASCE)CO.1943-7862.0000575.

Makridakis, Spyros, A. Andersen, Robert Carbone, Robert Fildes, Michele Hibon, Rudolf Lewandowski, Joseph Newton, Emanuel Parzen, and Robert Winkler. 1982. “The Accuracy of Extrapolation (time Series) Methods: Results of a Forecasting Competition.” Journal of Forecasting 1 (2): 111–53. doi:10.1002/for.3980010202.

Makridakis, Spyros, Chris Chatfield, Michèle Hibon, Michael Lawrence, Terence Mills, Keith Ord, and LeRoy F. Simmons. 1993. “The M2-Competition: A Real-Time Judgmentally Based Forecasting Study.” International Journal of Forecasting 9 (1): 5–22. doi:10.1016/0169-2070(93)90044-N.

Makridakis, Spyros, and Michèle Hibon. 2000. “The M3-Competition: Results, Conclusions and Implications.” International Journal of Forecasting, The M3- Competition, 16 (4): 451–76. doi:10.1016/S0169-2070(00)00057-1.

McCourt, R. S. 2004. Parking Generation. 3rd ed. https://trid.trb.org/view.aspx?id=754864.

Mead, Kenneth. 2004. The Rating and Evaluation of New Starts Transit Systems. Washington DC.

Memphis Business Journal Staff. 2009. “MATA Offers Free Trolley Rides Starting Wednesday.” Memphis Business Journal, April 19.

Merritt, H. Wm. 1993. “Reflections on the New Systems Study Project.” In Automated People Movers IV: Enhancing Values in Major Activity Centers : Proceedings of the Fourth International Conference, Irving, Texas, March 18, edited by William J. Sproule, Edward S. Neumann, and Murthy V. A. Bondada, 35–59. New York: Amer Society of Civil Engineers.

Metropolitan Atlanta Rapid Transit Authority. 2005. “History of MARTA: 1970-1979.” About MARTA: History. February 4. https://web.archive.org/web/20050204222807/http://itsmarta.com/about/history02.htm.

Miami-Dade County. 2015. “Miami-Dade Transit History.” Miamidade.gov. March 13. http://www.miamidade.gov/transit/history.asp.

Morrison, Neill. 1984. “The Accuracy of Quantity Surveyors’ Cost Estimating.” Construction Management and Economics 2 (1): 57–75. doi:10.1080/01446198400000006.

Murphy, Allan H. 1993. “What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting.” Weather and Forecasting 8 (2): 281–93. doi:10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Murphy, Allan H., and Robert L. Winkler. 1977. “Reliability of Subjective Probability Forecasts of Precipitation and Temperature.” Journal of the Royal Statistical Society. Series C (Applied Statistics) 26 (1): 41–47. doi:10.2307/2346866.

Næss, Petter, Bent Flyvbjerg, and Søren Buhl. 2006. “Do Road Planners Produce More ‘Honest Numbers’ than Rail Planners? An Analysis of Accuracy in Road-traffic Forecasts in Cities versus Peripheral Regions.” Transport Reviews 26 (5): 537–55. doi:10.1080/01441640500532005.

Niagara Frontier Transportation Authority. 2010. “History of Metro.” NFTA-Metro. http://metro.nfta.com/About/History.aspx.

Nyhan, Brendan, and Jason Reifler. 2010. “When Corrections Fail: The Persistence of Political Misperceptions.” Political Behavior 32 (2): 303–30. doi:10.1007/s11109-010-9112-2.

Nyhan, Brendan, Jason Reifler, Sean Richey, and Gary L. Freed. 2014. “Effective Messages in Vaccine Promotion: A Randomized Trial.” Pediatrics 133 (4): e835–42. doi:10.1542/peds.2013-2365.

Office of Inspector General. 2007. “Review of Congressional Earmarks within Department of Transportation Programs.” AV-2007-066. Washington DC: United States Department of Transportation. https://www.oig.dot.gov/sites/default/files/Congressial_Earmarks-_AV-2007-66----508_Compliant.pdf.

Office of Technology Assessment. 1975. Automated Guideway Transit : An Assessment of PRT and Other New Systems, Including Supporting Panel Reports. Washington DC: U.S. Government Printing Office.

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

Pickrell, Don H. 1989. “Urban Rail Transit Projects: Forecast versus Actual Ridership and Costs. Final Report.” Urban Mass Transit Administration. http://trid.trb.org/view.aspx?id=299240.

———. 1992. “A Desire Named Streetcar Fantasy and Fact in Rail Transit Planning.” Journal of the American Planning Association 58 (2): 158–76. doi:10.1080/01944369208975791.

Plotch, Philip Mark. 2015. “What’s Taking So Long? Identifying the Underlying Causes of Delays in Planning Transportation Megaprojects in the United States.” Journal of Planning Literature, January, 0885412214566116. doi:10.1177/0885412214566116.

Poe, Edgar Allan. 1836. “Maelzel’s Chess-Player.” Southern Literary Messenger 2 (5): 318–26.

Richmond, Jonathan. 2005. Transport Of Delight. Akron, Ohio: University Of Akron Press.

Rogers, Everett M. 1962. Diffusion of Innovations. 1st edition. The Free Press of Glencoe.

———. 2003. Diffusion of Innovations. 5th edition. New York: Free Press.

Rowland, Henry J. 1981. “The Causes and Effects of Change Orders on the Construction Process.” Master’s thesis. Georgia Institute of Technology.

Rowling, J. K. 1999. Harry Potter And The Prisoner Of Azkaban. 1st edition. New York: Scholastic.

San Diego Association of Governments. 1985. San Diego Trolley : The First Three Years. Washington DC: Urban Mass Transportation Administration. http://archive.org/details/sandiegotrolleyf00unit.

Schmitt, David. 2016a. "A Transit Forecasting Accuracy Database: Beginning to Enjoy the 'Outside View.'" In TRB 95th Annual Meeting Compendium of Papers. Washington DC: Transportation Research Board. https://trid.trb.org/view.aspx?id=1392575.

———. 2016b. "A Transit Forecasting Accuracy Database: Beginning to Enjoy the 'Outside View.'" In . https://trid.trb.org/view.aspx?id=1392575.

Schrag, Zachary M. 2008. The Great Society Subway. Reprint edition. Baltimore: Johns Hopkins University Press.

Short, Jack, and Andreas Kopp. 2005. "Transport Infrastructure: Investment and Planning. Policy and Research Aspects." Transport Policy 12 (4): 360–67. doi:10.1016/j.tranpol.2005.04.003.

Siemiatycki, Matti. 2009. "Academics and Auditors Comparing Perspectives on Transportation Project Cost Overruns." Journal of Planning Education and Research 29 (2): 142–56. doi:10.1177/0739456X09348798.

Silver, Nate. 2015. The Signal and the Noise: Why So Many Predictions Fail--but Some Don't. New York: Penguin Books.

Skamris, Mette, and Bent Flyvbjerg. 1996. "Accuracy of Traffic Forecasts and Cost Estimates on Large Transportation Projects." Transportation Research Record: Journal of the Transportation Research Board 1518 (January): 65–69. doi:10.3141/1518-12.

Smith, Jason Scott. 2006. Building New Deal Liberalism: The Political Economy of Public Works, 1933-1956. Cambridge University Press.

Smith, Stanley K. 1987. “Tests of Forecast Accuracy and Bias for County Population Projections.” Journal of the American Statistical Association 82 (400): 991–1003. doi:10.1080/01621459.1987.10478528.

Smith, Stanley K., and Terry Sincich. 1992. “Evaluating the Forecast Accuracy and Bias of Alternative Population Projections for States.” International Journal of Forecasting, Population Forecasting, 8 (3): 495–508. doi:10.1016/0169-2070(92)90060-M.

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

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

Tetlock, Philip E. 2009. Expert Political Judgment: How Good Is It? How Can We Know? Kindle Edition. Princeton University Press.

Theil, Henri. 1966. Applied Economic Forecasting. Chicago: Randy-McNally.

Tofallis, Chris. 2015. “A Better Measure of Relative Prediction Accuracy for Model Selection and Model Estimation.” Journal of the Operational Research Society 66 (8): 1352–62. doi:10.1057/jors.2014.103.

Trost, Steven M., and Garold D. Oberlender. 2003. “Predicting Accuracy of Early Cost Estimates Using Factor Analysis and Multivariate Regression.” Journal of Construction Engineering and Management 129 (2): 198–204. doi:10.1061/(ASCE)0733-9364(2003)129:2(198).

Tucker, A. W. 1983. “The Mathematics of Tucker: A Sampler.” The Two-Year College Mathematics Journal 14 (3): 228–32. doi:10.2307/3027092.

United States Census Bureau. 2000. “2000 Census of Population and Housing.” http://www.census.gov/prod/www/decennial.html.

United States General Accounting Office. 1980. “Better Justifications Needed for Automated People Mover Demonstration Projects.” CED-80-98. Washington DC: United States General Accounting Office.

Urban Mass Transportation Administration. 1976. “Federal Policy on Assistance for Major Urban Mass Transportation Investments.” Federal Register 41 (185): 41512–14.

———. 1978. “Policy Toward Rail Transit.” Federal Register 43 (45): 9428–30.

———. 1984. “Major Urban Mass Transportation Capital Investments Statement of Policy.” Federal Register 49 (98): 21284–91.

———. 1987. “The Status of the Nation’s Local Mass Transportation: Performance and Conditions, Report to Congress.” United States Department of Transportation.

———. 1990. “Report on Funding Levels and Allocations of Funds: Report of the Secretary of Transportation to the United States Congress Pursuant to Section 3(j) of the Urban Mass Transportation Act of 1964, as Amended.” Washington DC: U.S. Department of Transportation.

———. 1991. “Report on Funding Levels and Allocations of Funds: Report of the Secretary of Transportation to the United States Congress Pursuant to Section 3(j) of the Urban Mass Transportation Act of 1964, as Amended.” Washington DC: U.S. Department of Transportation.

U.S. Energy Information Administration. 2016. “State Energy Data System (SEDS): 1960-2014 (complete).” http://www.eia.gov/state/seds/sep_prices/total/csv/pr_all.csv.

———. 2016. “Energy Units - Energy Explained, Your Guide To Understanding Energy.” Accessed November 10. http://www.eia.gov/energyexplained/index.cfm?page=about_energy_units.

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

Walker, Richard, Gray Brechin, Gabe Milner, Harvey Smith, Susan Ives, Lisa Thompson, John Stehlin, et al. 2016. “New Deal Category: Mass Transportation.” Living New Deal. https://livingnewdeal.org/new-deal-categories/infrastructure/mass-transportation/.

Washington Metropolitan Area Transit Authority. 2016. “Metro Facts 2016.” http://www.wmata.com/about_metro/docs/Metro%20Facts%202016.pdf?

Weiner, Edward. 1999. Urban Transportation Planning in the United States: An Historical Overview. 2 edition. Westport, Conn: Praeger.

Welde, Moten, and James Odeck. 2011. “Do Planners Get It Right? The Accuracy of Travel Demand Forecasting in Norway.” European Journal of Transport and Infrastructure Research 11 (1): 80–95.

White, Helen R. 1954. “Empirical Study of the Accuracy of Selected Methods of Projecting State Populations*.” Journal of the American Statistical Association 49 (267): 480–98. doi:10.1080/01621459.1954.10483516.

Wilson, James Q, and George L. Kelling. 1982. “The Police and Neighborhood Safety: Broken Windows.” Atlantic Monthly 127 (2).

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