Routledge
Taylor & Francis Group

Transport Reviews

A Transnational Transdisciplinary Journal

ISSN: 0144-1647 (Print) 1464-5327 (Online) Journal homepage: http://www.tandfonline.com/loi/ttrv20

Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review

Morten Skou Nicolaisen & Patrick Arthur Driscoll

To cite this article: Morten Skou Nicolaisen & Patrick Arthur Driscoll (2014) Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review, Transport Reviews, 34:4, 540-557, DOI: 10.1080/01441647.2014.926428

To link to this article: http://dx.doi.org/10.1080/01441647.2014.926428

Transport Reviews, 2014
Vol. 34, No. 4, 540–557, http://dx.doi.org/10.1080/01441647.2014.926428

Ex-Post Evaluations of Demand Forecast Accuracy: A Literature Review

MORTEN SKOU NICOLAISEN*§ AND
PATRICK ARTHUR DRISCOLL**

**Department of Development and Planning, Aalborg University, Skibbrogade 5, Aalborg 9000, Denmark; **Department of Development and Planning, Aalborg University Copenhagen, AC Meyers Vænge 15, Copenhagen 2450, Denmark*

(*Received 9 June 2013; revised 12 May 2014; accepted 17 May 2014*)

Abstract *Travel demand forecasts play a crucial role in the preparation of decision support to policy-makers in the field of transport planning. The results feed directly into impact appraisals such as cost–benefit analyses and environmental impact assessments, which are mandatory for large public works projects in many countries. Over the last few decades, there has been increasing attention given to the lack of demand forecast accuracy. However, since data availability for comprehensive* ex-post *appraisals is problematic, such studies are still relatively rare. This study presents a review of the largest* ex-post *studies of demand forecast accuracy for transport infrastructure projects. The focus is threefold: to provide an overview of observed levels of demand forecast inaccuracy, to highlight key contextual and methodological differences between studies and to highlight key focus areas for future research in this field. The results show that inaccuracy remains problematic for road, rail and toll projects alike, but also how the lack of methodological clarity and consistency calls for a careful interpretation of these results. Mandatory, systematic* ex-post *evaluation programmes are suggested as a necessary tool to improve decision support, as data availability for* ex-post *studies is often remarkably poor even for internal audits.*

1. Introduction

Decision-making in the field of transport infrastructure planning relies extensively on various types of impact assessments. The most common contemporary appraisal method is cost–benefit analysis (CBA) (Hayashi & Morisugi, 2000; Mackie, 2010; Odgaard, Kelly, & Laird, 2005). A key aspect in this method of appraisal is the reliance on accurate forecasts for a range of impact categories, of which the most important for transport infrastructure projects are primarily construction costs and travel demand. The former is usually the biggest expense item while the latter is used to estimate total travel time savings, which is by far the most dominant economic benefit in conventional appraisal (Banister, 2008; Nicolaisen, 2012). The accuracy of cost estimates and demand forecasts is

§Corresponding author. Email: mortenn@plan.aau.dk

thus of great importance, as these predictions are some of the most crucial inputs to *ex-ante* appraisals of transport infrastructure. It should be noted that the economic performance of projects is not always the decisive political priority (Sager & Ravlum, 2005), just as conventional evaluation methods omit many indirect economic benefits of transport infrastructure investments (Banister & Thurstain-Goodwin, 2011). Nonetheless, several countries operate with a soft threshold value for economic performance that projects are benchmarked against, just as evaluation of many important effects depends on accurate demand forecasts even if these are included in a non-monetised form (e.g. noise, CO_2 emissions, accidents, etc.).

The forecasting potential of model-based decision support has long been subject to critical scrutiny, but typically the main academic focus has been directed at investigating levels of uncertainty related to model specification issues, calibration procedures or data acquisition for necessary input variables (Ascher, 1981; Clarke, Dix, & Jones, 1981; De Jong et al., 2007; De Jongh, 1998; Rasouli & Timmermans, 2012; Walker et al., 2003; Zhao & Kockelman, 2002). These are certainly important in and of themselves, but they are typically related to the issue of uncertainty rather than inaccuracy, where uncertainty is defined as "any departure from the unachievable ideal of complete determinism" (Walker et al., 2003, p. 8). Meanwhile, inaccuracy is defined as the actual deviation between predicted and observed values, which is only possible to measure after the true value is known. It is observed inaccuracy rather than estimated uncertainty that is of interest in the present study.

In contrast to the large body of literature on *ex-ante* appraisals to inform decision-makers of estimated uncertainty levels, there is a relatively sparse body of literature that deals with *ex-post* appraisals to inform decision-makers about observed inaccuracy levels (Van Wee, 2007). The issue is less problematic for cost estimates than for demand forecasts, since empirical studies of cost estimates are generally in ample supply, all of which seem to indicate a tendency for systematic cost overruns (Siemiatycki, 2009). This tendency is not restricted to planning of transport infrastructure projects (see, e.g. Hall, 1980; Kharbanda & Stallworthy, 1983), but due to the influential studies by Flyvbjerg (2007) and colleagues it has probably received increased attention here compared to other fields of planning. However, large-scale *ex-post* appraisals of demand forecasts remain relatively rare, both in transport infrastructure planning and elsewhere (Hartgen, 2013; Nicolaisen, 2012; Van Wee, 2007). Data acquisition for such studies is cumbersome, and while there has been a surge of interest over the last decade, the number of comparable studies is still relatively modest. The purpose of the present review is to bring together the main existing contributions and provide a comprehensive overview of the observed degree of demand forecast inaccuracy for transport infrastructure projects. This includes both the mean differences between forecasted and observed demand (i.e. whether a bias is present) as well as the standard deviation of the distribution (i.e. the overall level of precision). In addition, the methodological and contextual backgrounds of the studies will be examined to assess their comparability, the likely causes of inaccuracy, and fruitful areas of further research. Finally, a set of recommendations for improving future evaluation practice is provided. For a well-structured review of related issues for costs estimates, see Siemiatycki (2009).

2. Prior Reviews

The authors of the present paper are aware of only a few studies that have previously tried to provide a similar overview of findings in this field, none of which satisfies the purpose of the present review. Van Wee (2007) provides a joint review of the quality of cost and demand forecasts for large infrastructure projects, but the review only summarises the findings of two *ex-post* appraisals of demand forecasts, both of which shall also be covered here in more detail. The limited number of studies included in Van Wee's contribution is partly due to differences in scope, as the author also covers the accuracy of cost estimates. However, the very limited items included for review brings into question the robustness of the findings. Brinkman (2003) and Nicolaisen (2012) both offer more detailed overviews of prior studies of demand forecast accuracy. However, in both cases these serve mainly as context for other inquiries and omit several studies that would be equally relevant for the purposes of the present review.

In general, there is a significant lack of data availability for undertaking comprehensive *ex-post* appraisals, which limits the options for conducting such studies. Parthasarathi and Levinson (2010, p. 438) describe how "the data collection efforts were much more laborious and time consuming than anticipated. The unavailability of data in electronic format, lack of proper documentation, poor record keeping and data archiving procedures complicated the data collection process and subsequent analysis". Similar comments are found in more or less all *ex-post* appraisals that the authors of the present review have analysed. Apart from the obvious problem of making the acquisition of sufficient data for the studies difficult, the issue presents a more complicated problem in the form of availability bias (Tversky & Kahneman, 1973). Extreme cases risk being over-represented in study samples when data availability is poor, since projects that deviate the most from expectations are likely to be subject to the most intense critical scrutiny and thus provide more readily available data sources (Nicolaisen, 2012; Siemiatycki, 2009; Skaburskis & Teitz, 2003).

Given the limited availability of project data, the small number of prior studies and the lack of comprehensive reviews, the main motivations for the present paper are to provide a systematic overview of findings, compare the approaches used and provide recommendations for future research in the field.

3. Methodology

In order to identify the studies that would be selected for the review, a number of means were employed. First, iterative searches within Google Scholar, Web of Science and Scopus were conducted over a period of 24 months,[1] using the following search terms: '*ex-post*', '*ex-ante*/*ex-post*', 'post-opening evaluation', 'project appraisal', 'post-implementation evaluation' and 'transport project evaluation/audit'. Second, the bibliographies of the more frequently cited studies were cross-referenced against one another, to ensure completeness and identify blind spots. Third, the websites of the road and rail authorities and national audit offices in North America, Europe, Australia and New Zealand were mined for studies that evaluated either cost or demand forecast data for infrastructure investment.

The selection of studies included in the present review has been, out of necessity, quite pragmatic, since data availability is the most limiting factor of *ex-post* appraisals and has also been the dominant selection criteria in all reviewed studies. The studies included in the present review have been selected based on the following criteria:

(1) The study employs a sizeable database of completed rail and/or road projects (rail: $N \geq 10$, road $N \geq 50$).[2]
(2) The studies must provide either aggregate distributions of demand forecast inaccuracy or figures for individual projects that allow such distributions to be established.
(3) The studies must specify the sources of forecasts and observed values for travel demand.

Many studies have been excluded as a result of criterion 1, since they have focused on individual projects or included very small samples. The main reason for the sample size criterion is a considerable risk of availability bias and representativeness in smaller studies (Nicolaisen, 2012; Tversky & Kahneman, 1973). The small threshold for rail studies means that this is in itself an insufficient criterion for eliminating this risk, but in these cases the sample typically represents all projects completed in a given period. Interested readers may consult Brinkman (2003) for an overview of many of the studies excluded here.

A total of 12 studies that fulfil the above criteria were identified (Table 1). First, for each of these studies a distribution of demand forecast inaccuracy has been constructed by converting reported values to a standardised unit (see the next section for specification of this unit). The purpose is to enable a meaningful comparison of the reported levels of inaccuracy across studies. Second, a comparison of the key methodological differences has been conducted. Focus has been placed upon the reference points for forecast and observed values and possible adjustment measures, with the purpose of establishing how comparable the reported values of inaccuracy are. Third, the stated explanations for the reported levels of inaccuracy have been compared along with the available evidence presented by the authors in support of them. The main purpose is to identify the most commonly referenced causal explanations put forth by the authors.

Table 1. Prior studies of demand forecast inaccuracy included for review in the present paper

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

4. Observed Levels of Inaccuracy

The first aspect that is reviewed is the distribution of demand forecast inaccuracy observed in *ex-post* evaluations. The employed measures of inaccuracy differ between studies and require conversion to a standardised unit. In line with Flyvbjerg et al. (2006, p. 3) "common practice is followed here and the inaccuracy of a traffic forecast is defined as actual minus forecasted traffic in the percentage of forecasted traffic"[3]:

$$I = \frac{O - F}{F}.$$

This provides a simple measure of inaccuracy as the relative deviation between observed and expected values. Perfect accuracy is indicated by a measure of zero. Negative values indicate less demand than expected (i.e. demand has been overestimated), while positive values indicate more demand than expected (i.e. demand has been underestimated). A more sophisticated measure of inaccuracy would be desirable, since the chosen measurement is a point estimate that is sensitive to fluctuations in the opening year. However, data availability rarely allows for more than such simple measure to be established for completed projects. In addition, the measurement has a clear lower bound but no upper bound, meaning that a heavy right-tailed distribution rather than a true normal distribution could be an issue. Salling and Leleur (2006) suggest an Erlang distribution as a better fit for cost estimates, but since the results in of most studies of demand forecast inaccuracy approximate a true normal distribution, this distribution has been chosen as a suitable fit for the purposes of the present review.

Table 2 provides an overview of the mean and standard deviation of demand forecast inaccuracy across the range of reviewed *ex-post* evaluations. A few patterns of general interest should be noted here. First, the mean inaccuracy for road projects is typically positive, indicating that more demand than expected materialises after projects have been completed. Most results lie within the span of 3–11% additional traffic compared to forecasts, although two studies deviate from this pattern (Mackinder & Evans, 1981; Welde & Odeck, 2011). Both of these studies have relatively low sample sizes when compared to the remaining studies of road projects, which might explain some of this deviation. In addition, the economic conditions have been relatively unusual in the case areas at the time the studies were made (oil crises in the 1970s, booming Norwegian economy in the 2000s). Out of seven total road studies, the observed positive bias is statistically significant at $p \leq .05$ in four of them. Another two also report positive biases, although these are not statistically significant even at $p < .10$. Mackinder and Evans (1981) provide insufficient data for an analysis of statistical significance. It should be noted that significance testing for such a rough indicator is mostly to serve as a consistency check, and from the results it appears that there is indeed a general tendency to underestimate travel demand for road projects.

Second, the mean inaccuracy for rail projects is negative in all studies, indicating that less demand than expected materialised after projects have been completed. Most results lie within the span of 16–44% less patronage compared to forecasts, but one study reports a considerably lower mean than others (Pickrell, 1990). As was the case for road projects, the most extreme mean value is found in the study with the smallest sample. However, unlike the situation for road

Table 2. Comparison of means and standard deviations for observed demand forecast inaccuracy

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

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

projects, the mean deviates substantially even among studies with relatively large samples. Despite this large variation in mean values, the negative bias is statistically significant at the 0.05 level in all but one rail study. There appears to be a general tendency of severely overestimating travel demand for rail projects.

Third, the mean inaccuracy for toll projects is typically also negative, indicating that less demand than expected materialises after projects have been completed. Very few large-scale studies provide empirical evidence for this type of project, and the magnitude of bias differs quite substantially between the two studies that have been available for the present review. Bain (2009) reports a systematic tendency towards overestimation in demand forecasts for toll projects at a global scale, which is supported by findings in a smaller study of Australian toll road projects (Li & Hensher, 2010) that did not meet the sample size requirement for the present review. In contrast, the results from Welde and Odeck (2011) indicate that toll road forecasts are generally quite accurate in Norway. However, demand on non-tolled Norwegian road projects was severely underestimated in this period. It is therefore possible that the relative attractiveness of Norwegian toll roads has also been optimistic, but that this optimism has been countered by a general surge in travel demand on all roads.

Fourth, for all three project types (road, rail and toll) the inaccuracy measurement has considerable variation, indicating that the general trends observed from the mean values should not be used as more than crude indicators, even in cases where the trends are statistically significant. This is hopefully self-evident, but as there have been examples of rather naïve interpretations of the results (e.g. Cox & Moore, 2012), the issue will be given some attention in the discussion section.

Fifth, simple spatial or temporal categorisations do not appear to explain much of the observed inaccuracy. The geographical location of projects is not a good indicator of inaccuracy for any project type, neither in across studies nor in individual studies spanning multiple countries. For road projects there appears to be little or no improvement in accuracy over time, while some improvement for rail projects can be identified. The lack of studies on toll road projects and the limited project identification offered in the existing studies make it difficult to gauge any possible impact of spatial or temporal categories for this project type.

Overall, potential bias seems to be a noticeable problem for road projects, but a more considerable problem for rail and toll projects. However, even in cases where bias does not appear to be problematic, the relatively large imprecision across the range of studies presents a challenge for the use of travel demand forecasts as decision support.

5. Comparability of Projects

The second aspect that is reviewed is how comparable the results of different studies are based on the specification of demand forecast inaccuracy that has been employed. Multiple forecasts are produced during the planning phase of large infrastructure projects, and project designs often change during construction to accommodate new requirements. This makes it important to specify which forecast to use as a base of comparison, as it will often be possible to identify both underestimated and overestimated forecasts for any single project (Olsson, Krane, Rolstadås, & Veiseth, 2010).

The first column of Table 3 displays the source of forecasts and reference point for the forecast target year for each study. As can be seen, most studies refer to the CBA, the environmental impact assessment (EIA) or the traffic impact study (TIS) available at the time of decision to build. Some studies use these interchangeably

Table 3. Comparison of the sources used for demand forecasts, reference points for the forecast target year, and whether an adjustment measure has been employed in case of mismatching reference points for forecast and observed values

Author(s)	Source of forecast	Target year	Reference adjusted
Mackinder and Evans (1981)	TIS	Opening year	✓
NAO (1988)	TIS	Design year	✓
Pickrell (1990)	Varies	Varies	
Flyvbjerg et al. (2006)	Varies	Varies	
DoT (2007)	EIA	Opening year	✓
DoT (2008)	EIA	Opening year	✓
Bain (2009)	Financial close	Opening year	N/A
Button et al. (2010)	N/A	Varies	N/A
Parthasarathi and Levinson (2010)	TIS/EIA	Opening year	N/A
HA (2011)	CBA/EIA/TIS	Opening year	✓
Welde and Odeck (2011)	Parliamentary bill	Opening year	N/A
Nicolaisen (2012)	CBA/EIA/TIS	Opening year	✓

Note: CBA, cost–benefit analysis; EIA, environmental impact assessment; TIS, traffic impact study.

due to limited availability of consistent documentation, but as the forecast from the TIS is often reused in CBA and EIA documents, this issue is not considered a problem for comparing demand forecast inaccuracy. The same goes for the parliamentary bills used by Welde and Odeck (2011), as these usually refer to the values from the EIA. With the exception of HA (2011), none of the studies specify how they treat demand forecasts for different growth scenarios in these sources. It is assumed that a middle growth scenario has therefore been assumed in the remaining studies, and the results from HA (2011) have thus been interpolated from high- and low-growth scenarios to match. Three studies employ inconsistent sources of forecasts (Button et al., 2010; Flyvbjerg et al., 2006; Pickrell, 1990), which makes it difficult to compare the results of these directly with other studies. This is highly problematic, as the reported levels of demand forecast inaccuracy for rail projects are very different in these three studies, but where this may stem from using different sources of forecasts. In addition, Flyvbjerg et al. (2006) and Button et al. (2010) mainly pool results from previous studies with inconsistent sampling approaches, making it difficult to assess if the sampling approaches in these two studies are internally consistent.

The second column in Table 3 indicates the target year for which forecasts have been made. Most studies use the opening year as the forecast target year. The exceptions include the same three studies that also employ inconsistent sources as forecasts (Button et al., 2010; Flyvbjerg et al., 2006; Pickrell, 1990), likely due to the same reasons (i.e. pooling results of previous studies with different sampling approaches). This complicates the possibility of comparing the results of these studies directly even further. In addition to these three studies, the UK NAO (1988) uses the design year rather than the opening year as a reference point, meaning that forecasts are constructed for a system that is expected to have been in operation for 15 years. This is less problematic, as the target year is still internally consistent and the difference in methodology can thus at least be accounted for when comparing the results of the study with those from other studies.

The final column in Table 3 indicates whether a measure of adjustment has been applied in cases where the reference points for forecast and observed values do not match. Most forecasts are developed using a target year that rarely matches the actual opening year, either because projects are delayed or because the target year is intended to reflect a hypothetical opening year that does not necessarily correspond to an actual project schedule. For this reason, an adjusted measure is often preferred to reflect the most likely forecast for the actual opening year (e.g. if the forecast target year is 1997 but the project opens in 2004, the values in the forecast will be adjusted to 2004 based on the annual growth rates available at the time of constructing the forecast). However, some studies do not mention if or how such adjustments have been made, while others have omitted it deliberately. Pickrell (1990) sometimes compares forecasts with projects that are only partially completed, and some of the results pooled in Flyvbjerg et al. (2006) are derived from previous studies that also compare forecasts with partially completed projects or where there is a gap of several years between the reference points for forecasted and observed values.

Table 4 displays the unit of observations used to construct the inaccuracy measurements in each study. For road projects, using annual average daily traffic on either main links or across one or more screenlines is fairly standard. Both items are of interest, but the screenline analyses are often omitted due to

Table 4. The specified spatial unit of observation used to determine demand forecast accuracy

Author(s)	Unit of observation
Mackinder and Evans (1981)	AADT across screenline
NAO (1988)	AADT on road segments
Pickrell (1990)	Varies
Flyvbjerg et al. (2006)	Varies
DoT (2007)	Average weekday boardings
DoT (2008)	Average weekday boardings
Bain (2009)	N/A
Button et al. (2010)	N/A
Parthasarathi and Levinson (2010)	AADT on road segments
HA (2011)	AADT across screenline
Welde and Odeck (2011)	N/A
Nicolaisen (2012)	AADT on road segments, varies for rail projects

Note: AADT, annual average daily traffic.

lack of data availability. The studies of road projects that do include screenline analyses report less bias than those that only study individual road segments, indicating that the positive bias observed for most road projects might be a problem of predicting network distribution rather than travel demand. This indicates that new road capacity might attract more traffic from existing roads than expected, but so far this tendency has only been observed in a meta-analysis of completed projects conducted by the UK HA (2011). For rail projects, the units of observation are much less consistent. Some studies use boardings for the whole week, others use only use weekday boardings, and still others use different units for different projects depending on data availability. Table 4 thus points to yet another significant problem in comparing the result for rail projects directly due to inconsistent methodologies. For toll projects, none of the studies specify the units of observation in any detail, making it difficult to assess how comparable the results may be.

6. Explaining Inaccuracy

The reasons for the observed inaccuracy reported in the previous section are arguably one the most important issues to investigate if future forecasts and their application are to be improved. However, since the present article is concerned only with large-scale *ex-post* studies, the added breadth from such an approach has been associated with a corresponding lack of depth in the analysis of individual projects. Authors are typically hesitant to point out a specific source of inaccuracy as being dominant due to lack of convincing evidence. A limited set of explanation categories can be identified in the studies, based on the explanations reported by the authors of individual studies. An overview of the main explanations identified in the reviewed studies can be found in Table 5.

Auxiliary forecasts refer to inaccurate forecasts of exogenous variables that are likely to explain some of the observed demand forecast inaccuracy. As can be seen in Table 5, such explanations are offered in all of the *ex-post* studies that address the causes of inaccurate forecasts. Different economic growth trends than expected are the most frequently cited source of inaccuracy of input

Table 5. Comparison of the main explanations being offered for inaccuracy in demand forecasts

Author(s)	Auxiliary forecasts	Model specification	Design/operations	Network/land use	Politics/cognition
Mackinder and Evans (1981)	✓				
NAO (1988)	✓	✓			
Pickrell (1990)	✓	✓	✓		✓
Flyvbjerg et al. (2006)	✓	✓		✓	✓
DoT (2007)	✓		✓		
DoT (2008)	✓		✓	✓	
Bain (2009)	✓	✓		✓	✓
Button et al. (2010)	✓	✓			
Parthasarathi and Levinson (2010)	✓	✓		✓	✓
HA (2011)	✓	✓		✓	
Welde and Odeck (2011)	✓	✓			
Nicolaisen (2012)	✓	✓		✓	✓

variables, but car ownership and fuel prices are also commonly mentioned. Assumption drag, as it is described by Ascher (1981), appears to be one of the main reasons for the failure to provide accurate forecasts of key input variables. Some studies have provided quite convincing analyses of the likely impact of inaccurate input variables, such as Pickrell's (1990) comparison of estimated and actual development of exogenous variables for each project. Other studies rely on cruder assessment techniques, such as NAO's (1988) comparison of regional economic growth trends with national economic growth trends combined with informal feedback from local authorities.

Model specification is also commonly offered as an explanation of inaccuracy. However, very little documentation on what models have been used is provided in the studies, and most authors mention that access to such data is nearly impossible to acquire. Pickrell (1990, p. 29) notes that:

> errors arising from the way in which these models were applied, such as in the design and coding of transit networks, are extremely difficult to detect, yet they may be a major source of the ridership forecasting errors documented in this study.

Flyvbjerg et al. (2006) and Bain (2009) both find a lack of improvement in forecast accuracy over time in spite of better modelling tools having been developed, and therefore argue that model specification is likely not the most important area of improvement. Button et al. (2010) and Nicolaisen (2012) do find statistically significant improvements in forecast accuracy over time for rail projects, but do not associate this with increased model sophistication specifically. For road projects, a common neglect of induced demand effects is cited as a likely explanation for the observed bias towards underestimation by Welde and Odeck (2011), HA (2011) and Nicolaisen (2012). Given the lack of model documentation available for individual projects this is difficult to assess in detail, but several other studies support the claim that neglect of induced demand is or has been a common problem in many countries (Ladd, 2012; Litman, 2012; Mackie, 2010; Marte, 2003; MOTOS, 2007; Næss, Nicolaisen, & Strand, 2012; Nielsen & Fosgerau, 2005).

Design/operations refer to changes in project design or service operations that are likely to impact travel demand. The US DoT (2008) gives an example of a proposed light rail project, which expanded into a multi-modal corridor project with additional motorway capacity. This increased the costs of bridge crossings and parking facilities for the light rail project while reducing travel costs for a competing mode in the same corridor. Pickrell (1990) found that peak hour headway estimates for rail projects were lower than actual frequency for all projects examined, resulting in reduced capacity at the time of day when the projects were supposed to be the most competitive mode of transport. Such issues are only highlighted in studies of rail projects, which could be an indication that such projects are more susceptible to changing priorities due to the added complexity of service operations. This might partly explain why demand forecasts for such projects have a tendency of being more biased than road projects, as changes to headway and fare structures are relatively easy to implement compared to changes physical infrastructure.

Network/land use refers to changes in the built environment that are not specific to the project. For road projects, Flyvbjerg et al. (2006) found that changes in land-use development was one of the most common explanations for demand forecast inaccuracy, with 26% of projects experiencing this problem. Bain (2009), the US DoT (2008) and Nicolaisen (2012) each provide examples of changes in land use that had a severe impact on outturn demand for specific projects. Changes in land use can be particularly problematic for rail projects, if these are constructed as part of a land-use densification scheme that is never implemented. Nicolaisen (2012) provides an example of actual ridership being only 10% of demand forecasts for an individual station where surrounding development plans were cancelled. In addition to land use changes are issues of the actual transport network. Parthasarathi and Levinson (2010) did an analysis of the expected road networks used in the demand forecasts, and found that certain roads elsewhere in the network were still not completed 20 years later due to lack of funding, public opposition or changing priorities. As a result, the accuracy of both trip generation and trip distribution in such models suffers accordingly. Several studies indicate that this source of inaccuracy often results from the reuse of old data sets or forecasts (as they are expensive to update) and poor internal communication (especially between land-use planners and transport planners).

Politics/communication relates to the political, institutional or psychological sources of forecast inaccuracy. Several authors point out that model results often suffer from misinterpretation by people who lack the necessary understanding of how the forecasts have been produced. Parthasarathi and Levinson (2010, p. 441) conclude that "criticisms against the use of modelling in forecasting arises when, for example, results are used by policy-makers who lack an understanding of the process behind the numbers". Nicolaisen (2012) lists poor or absent communication of underlying assumptions as a problem for the majority of projects, which can undermine trust in the results and may lead to opportunistic use of forecasts. A more cynical interpretation of this source of inaccuracy is that such mistakes are probably intentional, and that a more fitting description would be *deliberate manipulation*. Flyvbjerg et al. (2006) argue that this is the most important source of forecast inaccuracy for rail projects in order to obtain political approval for capital-intensive projects, and the authors base their claim on statements collected from researchers and project managers. Bain (2009) provides a similar argument for toll projects. Pickrell (1990, pp. 29–30) describes how "the misinterpretation of [the models'] numerical outputs during the planning process" is likely an important source of demand forecast inaccuracy, although no indication is given for whether this is intentional or not.

In addition to the above explanations offered by the authors of the reviewed studies, Eliasson and Fosgerau (2013) have recently argued that the observed biases in demand forecast accuracy may be a result of selection bias in the decision process. According to this study, even if no bias is present in the evaluation of individual projects, a rational selection process based on expected feasibility is likely to result in a biased distribution when evaluating completed projects. The study provides no evidence to discredit other theories such as that of deliberate manipulation, but it does offer the possibility of a less cynical, although not necessarily less problematic, explanation for the observed bias in travel demand forecasts.

The multitude of explanations offered and the methodological diversity of the evidence provided in their support should be taken as an indication of the difficulty in assessing causal mechanisms of demand forecast inaccuracy. It seems clear that more detailed access to specific project data and performance is necessary if researchers are to conduct more elaborate analyses of these matters. Interestingly, all of the studies that highlight politics and cognition as important explanations for demand forecast inaccuracy have been conducted by academics. None of the internal audits have investigated such issues, nor does there appear to be a tradition of cross-referencing the academic studies that do. The authors of the present review therefore second Siemiatycki's assertion that (2009, p. 155) "greater interaction between auditors and academics could tap into the unique strengths of both groups, leading to research that contributes to an expanded understanding of the patterns, causes of, and cures for cost overruns". The same is likely true for demand forecast inaccuracy.

7. Discussion

From the available evidence, it is clear that demand forecast inaccuracy is problematic for all project types examined here (road, toll and rail). A tendency to underestimate demand for road projects results in capacity limits being met earlier than anticipated. For fixed links this will typically require an expensive capacity expansion, and for congested networks it means that expected travel time savings will likely be overestimated (HA, 2011; Mogridge, 1997). As an example, Næss et al. (2012) found that just 5% of additional demand could result in more than one-third of the expected benefits being lost. For toll and rail projects, the tendency to overestimate demand results in declining fare revenue and lower overall user benefits. However, there is some indication that forecasts are improving in accuracy, as studies of newer toll and rail projects report both lower bias and standard deviation than those of older projects. It is possible that prior studies themselves have improved the attention to forecast inaccuracy, and that planners have improved their efforts in response. In addition, there appears to be a switch in focus towards strengthening the role of public transport in many countries, resulting in planners having more experience with these types of projects as well as rising demand for public transport in general. Both of these effects have likely helped to make forecasts for rail projects more robust. For road projects, the phenomenon of peak car travel (Goodwin, 2012) might partially explain why newer road projects appear less likely to experience more traffic than expected.

In addition, some projects might provide other benefits in spite of the failure to meet demand targets, especially when considering non-transport-related impact categories (Banister & Thurstain-Goodwin, 2011). The values reported in Table 2 should therefore not be interpreted as a proxy for inaccuracy of overall impacts, since subsequent appraisal categories do not necessarily have a linear relationship with the measure of demand inaccuracy employed here. It is beyond the scope of the present review to investigate such issues in detail.

Apart from the observed biases in demand forecasts, the large variation in accuracy is in itself problematic. The majority of studies report standard deviations above 30%, and half the studies of rail projects report standard deviations above 50%. Even if the average forecasting bias had been close to zero for all studies, the general imprecision of forecasts shows that quantified decision support based on travel demand forecasts makes it difficult to compare projects directly. This is not to say that demand forecasts cannot provide useful decision support, but it highlights the necessity of clearly communicating the assumptions upon which such forecasts are based (Hartgen, 2013). Many countries are continuously seeking to improve their forecasting practice by developing increasingly sophisticated modelling tools, but if many of the causes of demand forecast inaccuracy lie outside the model domain, this approach will likely not improve the quality of information in any significant way. For example, if overly optimistic headway figures or fare levels is a major cause of overestimating ridership for rail projects, it might be more important to critically scrutinise the viability of these assumptions rather than improving the transport models used to assess the projected demand.

It should be noted that while it is possible to identify certain tendencies from the available evidence, comprehensive empirical investigations of demand forecast accuracy remain relatively rare. There are three major areas of improvements that would greatly add to the current body of knowledge in this field. First, one of the biggest limitations in the *ex-post* studies reviewed here is the focus on project-specific links, whereas demand forecasts are typically produced for a full transport network. This means that the observed inaccuracies could potentially be the result of failures to accurately assess network distribution rather than the overall demand. This shortcoming is not trivial, and while a few studies have tried to address the issue (HA, 2011; Mackinder & Evans, 1981; Nicolaisen, 2012; Parthasarathi & Levinson, 2010) it has only been covered very superficially due to limited data availability. Further research should seek to address this problem by incorporating a broader network analysis, possibly by using screenlines rather than individual links as a basis for comparison. However, judging from the experiences described in existing studies, such an approach would require a significant upgrade in mandatory data collection and public availability, since the costs of data collection would otherwise be unfeasibly high for individual research purposes.

Second, the explanations for forecast inaccuracy offered by authors are difficult to generalise since the evidence offered in support of the claims varies in nature and quality. The studies that make the greatest effort to address this aspect are rarely able to provide more than rough indications of causal mechanisms. An example of this is Pickrell (1990), where an in-depth examination of a small sample provides the most comprehensive inquiry into causes of forecast inaccuracy among the reviewed studies. Part of this examination involves testing the explanatory power of inaccurate input variables by adjusting forecasts in line with the observed *ex-post* values. However, as the author has no access to the actual transport model used to produce the original forecast, this test is limited to adjusting patronage figures based on changes in key input variables by using elasticity values from a conference paper. To put this in perspective, the American Public Transit Association (1991) reported elasticity values for demand and fare levels in the range of 0.12–0.86 in different US cities. With such a large variation, it is difficult to verify the credibility of the conclusion that only half the variation can be explained from changes in this type of key input variable, since this depends greatly on which elasticity values were used in the original forecast. Further research should seek to address this problem through in-depth case studies in line with Kain (1990), as the necessary data access makes this an impossible task for large-scale studies.

Third, the lack of availability for necessary data items is a general problem and probably the biggest limitation to advances in the field. For studies of cost estimate inaccuracy, Siemiatycki (2009, p. 152) found that "numerous scholarly studies of cost overruns in the transportation sector have identified inconsistent data access as a key limitation of their research findings". For studies of demand forecast inaccuracy this can be extended beyond scholarly studies, as even the internal auditing agencies are often unable to retrieve the necessary data items due to poor archiving protocols. As an example, the UK Highways Agency has conducted Post-Opening Project Evaluations for every completed major scheme (projects exceeding £10 million) since 2002, which is arguably one of the more comprehensive audits of decision support for transport infrastructure projects to date. Despite having binding archival requirements for *ex-ante* appraisal documents, the HA (2011) has been unable to supply the necessary documentation to assess demand forecast accuracy in detail for one-third of the projects. When institutions with mandatory archival procedures are unable to locate decision support documents internally, it stands to reason that cross-institutional access to such documents for researchers is also quite limited. Unless systematic archiving of necessary documentation becomes an institutionalised practice, the quality of *ex-post* evaluations is unlikely to improve greatly in the future. This is not a task that falls on researchers, but instead requires prioritisation of systematic *ex-post* evaluations by the responsible national and supranational transport and finance authorities.

Systematic *ex-post* project evaluation is particularly relevant for the emerging practice of reference class forecasting (Flyvbjerg, 2007; Salling & Bannister, 2009), which seeks to adjust forecasts based on what Kahneman (2011) refers to as the outside perspective. The basic purpose of this approach is sound, since it aims at controlling for undesirable optimism bias in the evaluation of new projects by comparing them to the performance distribution of completed projects. In order for this approach to be of much value, though, it is a requirement that suitable reference classes can be established. However, the findings in the present review show that the necessary data requirements for establishing reference classes are currently quite poor, and crude classifications such as 'road and 'rail differ greatly between individual studies. This makes it difficult to select an appropriate reference class when appraising a new project. As an example, Cox and Moore (2012) used a simplified version of the approach to adjust patronage figures for a planned surface-level rail project in accordance with the distribution of demand forecast inaccuracy for rail projects reported in Flyvbjerg et al. (2006). However, it is highly doubtful whether the new project is comparable to those in the employed reference class, as this includes older metro systems in third-world countries. This is true for cost and demand forecasts alike. The purpose of this example is not to argue against the use of reference class forecasting in general. It can be very useful in establishing a desirable cost uplift factor in aggregate budgeting for a pool of projects. However, at the moment it seems quite unsuitable for individual project assessments, where it risks diverting attention from important sources of uncertainty related to project-specific features. The reference classes currently established are much too crude for such purposes due to the large standard deviations observed in Table 1. Rather, the observed forecast inaccuracy should serve to highlight the dangers in giving too much emphasis to quantified model results without adequately assessing the assumptions upon which they are based. This is true for traditional demand forecasting, reference class forecasting and the findings of the present review alike.

8. Takeaway for Practice

This review concludes that poor accuracy of travel demand forecasts remain a considerable problem in transport planning. Inaccuracy is a problem both in the form of observed biases (non-zero means) as well as general imprecision (large standard deviations). Planners and decision-makers should thus be wary of putting too much emphasis on detailed assessments of the transport-related benefits, since the accuracy of the underlying forecasts rarely justifies such detail.

The most important source of inaccuracy for demand forecasts appears to be auxiliary forecasts of exogenous variables, where inaccurate forecasts of economic growth, car ownership and migration patterns propagate into demand forecasts. This importance is likely to increase, as many countries are currently experiencing a stagnation or decline in car traffic combined with an accelerated rate of urbanisation and densification of existing urbanised areas. Whether these trends are temporary or permanent will have a major impact on the demand for different modes of transport. In addition, long-term commitments to the development of land use, service operations and fare levels are important factors in ensuring more accurate demand forecasts for rail projects.

Model specification remains a potentially large source of demand forecast inaccuracy, but none of the studies included in the present review have had access to data that allows the impact of this to be evaluated in detail. In most instances, it is difficult to pinpoint the model that has been used to produce a forecast, let alone the specific data files that were used. In general, lack of data availability has been an important obstacle in all reviewed studies and should be a key focus area of future research and practice on the use of demand forecasts as decision support. The lack of data access makes it difficult to perform more elaborate statistical analyses on exogenous variables, cover larger network effects, evaluate demand over time, and track changes in land use, project design or service levels. It ought to be a key priority for funding institutions to implement better monitoring of project impacts, develop standardised archival procedures and make the data publically available, in order to facilitate more detailed studies of why travel demand forecasts are still often highly inaccurate.

As a closing statement it seems relevant to inquire whether it is at all possible to significantly improve demand forecast accuracy. The present review indicates that inaccuracy has remained problematic for several decades in spite of improved forecasting techniques. At the same time, transport projects and the policies they are supposed to serve are growing increasingly complex, which increases uncertainty and thereby also the challenge of providing accurate demand forecasts. Hartgen (2013) outlines two possible approaches for how planning research and practice can overcome this problem. One is labelled *hubris*, in which attempts are made to substantially improve model-based forecasting by monitoring performance, improving models and modifying the institutional arrangements to reduce optimism bias. The other is labelled *humility*, in which attempts are made to quantify uncertainty, recognise the lack of accuracy and deliberately reduce the impact of forecasts on decision-making. Either of these approaches is likely to reduce the problem of demand forecast inaccuracy, but common to both of them is that they require greatly improved monitoring of completed projects. Mandatory, systematic evaluation programmes appear to be the most important first step in improving decision support for transport planning.

Notes

1. This task has been conducted during an on-going research project, hence the extensive timeframe.
2. The reason for the smaller threshold for rail projects is that most countries construct vastly more road than rail projects, and the available population for sampling is thus much smaller. Large-scale rail studies simply do not exist.
3. *I*, inaccuracy; *O*, observation; *F*, forecast.

References

American Public Transit Association. (1991). *Fare elasticity and its application to forecasting transit demand*. Washington, DC: Author.

Ascher, W. (1981). The forecasting potential of complex models. *Policy Sciences*, *13*(3), 247–267.

Bain, R. (2009). Error and optimism bias in toll road traffic forecasts. *Transportation*, *36*(5), 469–482.

Banister, D. (2008). The sustainable mobility paradigm. *Transport Policy*, *15*, 73–80.

Banister, D., & Thurstain-Goodwin, M. (2011). Quantification of the non-transport benefits resulting from rail investment. *Journal of Transport Geography*, *19*, 212–223.

Brinkman, A. (2003). *The ethical challenges and professional responses of travel demand forecasters* (PhD thesis). University of California, Berkeley.

Button, K. J., Doh, S., Hardy, M. H., Yuan, J., & Xin, Z. (2010). The accuracy of transit system ridership forecasts and capital cost estimates. *International Journal of Transport Economics*, *37*(2), 155–168.

Clarke, M., Dix, M., & Jones, P. (1981). Error and uncertainty in travel surveys. *Transportation*, *10*(2), 105–126.

Cox, W., & Moore, A. (2012). The XpressWest High-speed rail line from Victorville to Las Vegas: A taxpayer risk analysis. *Reason Foundation*. Retrieved April 10, 2013, from http://reason.org/studies/show/the-xpresswest-high-speed-rail-line

De Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., & Hofman, F. (2007). Uncertainty in traffic forecasts: Literature review and new results for the Netherlands. *Transportation*, *34*(4), 375–395.

De Jongh, P. (1998). Uncertainty in EIA. In P. Warthern (Ed.), *Environmental impact assessment: Theory and practice* (New ed., pp. 62–83). London: Routledge.

Department of Transportation. (2007). *Contractor performance assessment report*. US Department of Transportation, Federal Transit Administration, Office of Planning and Environment.

Department of Transportation. (2008). *The predicted and actual impacts of the new starts projects 2007*. US Department of Transportation, Federal Transit Administration, Office of Planning and Environment.

Eliasson, J., & Fosgerau, M. (2013). Cost overruns and demand shortfalls — deception or selection? *Transportation Research Part B: Methodological*, *57*, 105–113. doi:10.1016/j.trb.2013.09.005

Flyvbjerg, B. (2007). *Megaproject policy and planning: Problems, causes, cures* (Doctoral dissertation). Aalborg University.

Flyvbjerg, B., Holm, M. K. S., & Buhl, S. L. (2006). Inaccuracy in traffic forecasts. *Transport Reviews*, *26*(1), 1–24.

Goodwin, P. (2012). *Peak travel, peak car and the future of mobility.* Presented at the international transport forum, Leipzig.

Hall, P. (1980). *Great planning disasters*. London: Weidenfeld & Nicholson.

Hartgen, D. T. (2013). Hubris or humility? Accuracy issues for the next 50 years of travel demand modeling. *Transportation*, *40*(6), 1133–1157. doi:10.1007/s11116-013-9497-y

Hayashi, Y., & Morisugi, H. (2000). International comparison of background concept and methodology of transportation project appraisal. *Transport Policy*, *7*(1), 73–88.

Highways Agency. (2011). *Post opening project evaluation of major schemes (2002 to 2009): Meta analysis* (Main Report). UK Highways Agency.

Kahneman, D. (2011). *Thinking, fast and slow* (1st ed.). New York: Farrar, Straus and Giroux.

Kain, J. F. (1990). Deception in Dallas. Strategic misrepresentation in rail transit promotion and evaluation. *Journal of the American Planning Association*, *56*(2), 184–196.

Kharbanda, O. P., & Stallworthy, E. A. (1983). *How to learn from project disasters. True-life stories with a moral for management*. Aldershot: Gower.

Ladd, B. (2012). You can't build your way out of congestion— or can you? *disP — The Planning Review*, *48*(3), 16–23.

Li, Z., & Hensher, D. (2010). Toll roads in Australia: An overview of characteristics and accuracy of demand forecasts. *Transport Reviews*, *30*(5), 541–569.

Litman, T. (2012). *Generated traffic and induced travel. Implications for transport planning*. Victoria: pnVictoria Transport Policy Institute.

Mackie, P. (2010). *Cost-benefit analysis in transport. A UK perspective*. Discussion papers from the international transport forum, Mexico.

Mackinder, I. E., & Evans, S. E. (1981). *The predictive accuracy of British transport studies in urban areas* (No. 699). Crowthorne: Transport and Road Research Laboratory.

Marte, G. (2003). Slow vehicle traffic is a more attractive alternative to fast vehicle traffic than public transport. *World Transport Policy & Practice*, *9*(2), 18–23.

Mogridge, M. J. H. (1997). The self-defeating nature of urban road capacity policy. *Transport Policy*, *4*(1), 5–23.

MOTOS. (2007). *Transport modelling: Towards operational standards in Europe* (Handbook No. MOTOS/M2.1/PU/v1.0). MOTOS project EU.

Næss, P., Nicolaisen, M. S., & Strand, A. (2012). Traffic Forecasts ignoring induced demand: A shaky fundament for cost-benefit analyses. *European Journal of Transport and Infrastructure Research*, *12*(3), 291–309.

National Audit Office. (1988). *Department of transport, Scottish development department and Welsh office: Road planning* (No. 688). London: Author.

Nicolaisen, M. S. (2012). *Forecasts: Fact or fiction? Uncertainty and inaccuracy in transport project evaluation* (PhD thesis). Aalborg University.

Nielsen, O. A., & Fosgerau, M. (2005). *Overvurderes tidsbenefit af vejprojekter?* Proceedings from the annual transport conference at Aalborg University, Aalborg.

Odgaard, T., Kelly, C., & Laird, J. (2005). *Current practice in project appraisal in Europe. Analysis of country reports* (No. Deliverable 1/Volume 1). European Commission EC-DG TREN.

Olsson, N. O. E., Krane, H. P., Rolstadås, A., & Veiseth, M. (2010). Influence of reference points in ex post evaluations of rail infrastructure projects. *Transport Policy*, *17*(4), 251–258. doi:10.1016/j.tranpol.2010.01.008

Parthasarathi, P., & Levinson, D. (2010). Post construction evaluation of traffic forecast accuracy. *Transport Policy*, *12*(6), 428–443.

Pickrell, D. H. (1990). *Urban rail transit projects: Forecast versus actual ridership and cost*. Washington, D.C.: Urban Mass Transportation Administration.

Rasouli, S., & Timmermans, H. (2012). Uncertainty in travel demand forecasting models: Literature review and research agenda. *Transportation Letters. The International Journal of Transportation Research*, *4*(1), 55–73.

Sager, T., & Ravlum, I. A. (2005). The political relevance of planners' analysis. The case of a parliamentary standing committee. *Planning Theory*, *4*(1), 33–65.

Salling, K. B., & Bannister, D. (2009). Assessment of large transport infrastructure projects: The CBA-DK model. *Transportation Research Part A*, *43*, 800–813.

Salling, K. B., & Leleur, S. (2006). *Assessment of transport infrastructure projects by the use of Monte Carlo simulation. The CBA-DK model.* Proceedings of the 2006 winter simulation conference, Monterey, 1537–1544.

Siemiatycki, M. (2009). Academics and auditors. Comparing perspectives on transportation project cost overruns. *Journal of Planning Education and Research*, *29*(2), 142–156. doi:10.1177/0739456X09348798

Skaburskis, A., & Teitz, M. B. (2003). Forecasts & outcomes. *Planning Theory and Practise*, *4*(4), 429–442.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Van Wee, B. (2007). Large infrastructure projects: A review of the quality of demand forecasts and cost estimations. *Environment and Planning B: Planning and Design*, *34*(4), 611–625.

Walker, W. E., Harremöes, P., Rotmans, J., Van Der Sluijs, J. P., Van Asselt, M. B. A., Janssen, J., . . . Krauss, M. P. (2003). Defining uncertainty a conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, *4*(1), 5–17.

Welde, M., & Odeck, J. (2011). Do planners get it right? The accuracy of travel demand forecasting in Norway. *European Journal of Transport and Infrastructure Research*, *1*(11), 80–95.

Zhao, Y., & Kockelman, K. (2002). The propagation of uncertainty through travel demand models: An exploratory analysis. *The Annals of Regional Science*, *36*(1), 145–163.