

# Traffic forecasts under uncertainty and capacity constraints

Anna Matas · Josep-Lluís Raymond · Adriana Ruiz

Published online: 24 February 2011  
© Springer Science+Business Media, LLC. 2011

**Abstract** Traffic forecasts provide essential input for the appraisal of transport investment projects. However, according to recent empirical evidence, long-term predictions are subject to high levels of uncertainty. This article quantifies uncertainty in traffic forecasts for the tolled motorway network in Spain. Uncertainty is quantified in the form of a confidence interval for the traffic forecast that includes both model uncertainty and input uncertainty. We apply a stochastic simulation process based on bootstrapping techniques. Furthermore, the article proposes a new methodology to account for capacity constraints in long-term traffic forecasts. Specifically, we suggest a dynamic model in which the speed of adjustment is related to the ratio between the actual traffic flow and the maximum capacity of the motorway. As an illustrative example, this methodology is applied to a specific public policy that consists of suppressing the toll on a certain motorway section before the concession expires.

**Keywords** Traffic forecast · Uncertainty · Toll motorways

## Introduction

Traffic forecasts provide essential input for the appraisal of transport investment projects and public policies. In spite of significant improvements to transport demand models over the past few decades, there are still high levels of uncertainty in long-term forecasts. For instance, a recent study by Flyvbjerg et al. (2006) concludes that accuracy in forecasting traffic flow has not improved over time. Given that project profitability is highly dependent on predicted traffic flow, uncertainty has to be quantified and accounted for in project evaluation.

---

A. Matas (✉) · A. Ruiz  
Departament d'Economia Aplicada, Universitat Autònoma de Barcelona, Barcelona, Spain  
e-mail: anna.matas@uab.es

J.-L. Raymond  
Departament d'Economia i Història Econòmica, Universitat Autònoma de Barcelona, Barcelona, Spain

This article quantifies uncertainty in traffic forecasts for the tolled motorway network in Spain. We estimate a demand model using a panel data set covering 67 tolled motorway sections between 1980 and 2008. Uncertainty is quantified in the form of a confidence interval for the traffic forecast that takes account of both the variance of the traffic forecast related to the stochastic character of the model (model uncertainty) and the uncertainty that underlies the future values of the exogenous variables (input uncertainty). Furthermore, as an illustrative example we apply this methodology to a specific public policy consisting of suppressing the toll on a certain motorway sections before the concession expires. In this case, the government has to compensate the private motorway concessionaire for the revenue forgone up to the end of the concession period. We present a point estimate for the present value of the forgone revenue, as if the result were certain, and then a set of confidence intervals at different levels of significance that account for the variance of the forecasting error.

The predictions are based on an aggregate demand equation, where traffic flow depends on the following variables: Gross Domestic Product, toll per kilometre, petrol price and a set of dummy variables that account for major changes in the road network. However, if maximum infrastructure capacity is not allowed for in the model, it may well be that predictions lie above this maximum value. To avoid this problem we should, ideally, estimate an integrated demand–supply system. However, as is often the case, we are not able to model the supply side of the system due to lack of data. Our article contributes to this issue by proposing a new functional form for the demand equation that accounts for the fact that the rate of growth of traffic flow diminishes as the volume approaches full capacity. Specifically, as detailed in “[The model](#)” section, we suggest a modified partial adjustment model with variable adjustment speed. In our case, and in terms of forecasting capacity, this proposal is preferable to the traditional logistic functional form with a saturation level equal to maximum capacity, given that we avoid the assumption that traffic follows an S-shaped growth curve.

## Literature review of uncertainty in traffic forecasting

Several recent studies confirm the inaccuracy of traffic predictions. Among them, the extensive work by Flyvbjerg et al. (2006) based on 210 transport infrastructure projects in 14 nations, 27 of which correspond to rail projects and the rest to road projects. They conclude that passenger forecasts for nine out of ten rail projects are overestimated, with an average overestimation of 106%. The authors suggest that there is a systematic positive bias in rail traffic forecasts. For road projects, forecasts are more accurate and balanced, although for 50% of the projects the difference between actual and forecasted traffic was more than  $\pm 20\%$ .<sup>1</sup> For both road and rail projects, the estimated standard deviation of the forecasting error is high, showing a high level of uncertainty and risk.

Bain (2009) presents the results from a study that analyses the performance of traffic forecasts for toll road traffic from a database including over 100 international toll road projects. The research confirms a large range of error in traffic forecasting and the existence of systematic optimism bias. On average, toll road forecasts overestimated first-year traffic by 20–30%.

---

<sup>1</sup> The authors suggest reference class forecasting as an alternative methodology. This proposal is detailed in Flyvbjerg (2008).

Using data on 14 toll motorway concessions in Spain, Vassallo and Baeza (2007) found that, on average, actual traffic during the first 3 years of operation was overestimated by approximately 35%. They conclude that there is a substantial optimism bias in the ramp-up period for toll motorway concessions in Spain.

The aforementioned studies suggest that the positive bias found for rail and toll motorways appears when there is a strong will for the approval of the project.

In spite of the significant errors present in traffic forecasting, uncertainty is often a neglected issue. Most of the predictions are presented as point estimates and the probability distribution of the outcome is forgotten about. The most common way to deal with uncertainty is to present alternative estimates based on different scenarios for the exogenous variables. However, this approach does not recognise all sources of uncertainty and, most importantly, does not provide the likelihood of each alternative forecast.

As stated by de Jong et al. (2007), the literature on quantifying uncertainty in traffic forecasting is fairly limited. The author reviews a considerable amount of the literature on that subject considering both the methodology employed and the results obtained. He distinguishes between input uncertainty, associated with the fact that future values of the exogenous variables are unknown, and model uncertainty, which includes random term uncertainty and coefficient uncertainty. Given that the 21 studies reviewed use different measures to express uncertainty and many of them do not present quantitative outcomes, providing an order of magnitude for uncertainty is difficult. De Jong suggests that input uncertainty is more important than model uncertainty; studies on input uncertainty or both input and model uncertainty obtain 95% confidence intervals for the mean value of traffic flow between  $\pm 18$  and  $\pm 33\%$ . The aforementioned paper also offers a methodology for quantifying uncertainty for a case study in The Netherlands.

The literature shows that quantifying forecast uncertainty and its causes is an area that deserves more attention. This article intends to contribute to this issue with new findings.

## The model

Given that the demand equation is estimated in order to predict future traffic flow, when specifying the equation we should take into account that as the volume of traffic increases, costs related to congestion emerge and the rate of traffic growth diminishes as traffic volume approaches maximum capacity of a road section. Toll motorways were introduced in the early 1970s on the road network in Spain. Nowadays, some of these motorways are close to their maximum capacity. This problem mainly affects those toll roads near urban areas and the main corridor along the Mediterranean coast, where it is difficult and costly to expand capacity. In these cases capacity constraints need to be considered when forecasting in order to avoid excessively optimistic results.<sup>2</sup>

Ideally, congestion costs and capacity constraints should be accounted for through a network assignment model, allowing a feedback between the various stages of the travel demand forecasting process.<sup>3</sup> However, frequently such a model is unavailable. As an

---

<sup>2</sup> The proposed functional form can be useful with other transport modes. The authors have applied it to forecasting air transport. With respect to this transport mode, Riddington (2006) concludes that the air traffic predictions in the United Kingdom are excessively optimistic and that no consideration is given to the existing restrictions on the capacity to handle this forecast demand.

<sup>3</sup> A sensible policy for motorway operators would be to increase tolls as demand approaches maximum capacity. Nonetheless, this is not a feasible policy in Spain as toll rates are set by law and are, therefore,

alternative approach, we suggest a functional form that can be considered as an implicit reduced form for the demand function. Specifically, we estimate a modified partial adjustment model, where the speed of adjustment is variable. The proposed equation can be derived as follows:

The static equation of the partial adjustment model takes the standard form and shows the logarithm of the equilibrium value of traffic  $Y_{it}^*$  on road section  $i$  in period  $t$  as a function of a set of variables  $X_{it}$ :

$$\ln Y_{it}^* = \alpha_i + \beta \ln X_{it} \quad (1)$$

The dynamic of the adjustment is modified by introducing a variable adjustment parameter,  $\lambda_{it}$ :

$$\Delta \ln Y_{it} = \ln Y_{it} - \ln Y_{it-1} = \lambda_{it} \cdot (\ln Y_{it}^* - \ln Y_{it-1}) + \varepsilon_{it} \quad (2)$$

We assume that the speed of adjustment decreases as traffic flow increases in the following terms. Let us define the quality level of the motorway,  $\tau$ , as a function of the traffic flow related to the maximum capacity of the infrastructure,  $Y_i^{\max}$ :

$$\tau_{it} = \frac{Y_i^{\max} - Y_{it-1}}{Y_i^{\max}} \quad (3)$$

Then, the adjustment parameter is assumed to be a function of  $\tau_{it}$ :

$$\lambda_{it} = \theta \left( \frac{Y_i^{\max} - Y_{it-1}}{Y_i^{\max}} \right) = \theta \cdot \tau_{it} \quad (4)$$

where  $\theta$  is a parameter that links the speed of adjustment and the level of use of the motorway section.

This functional form accounts for the fact that the rate of traffic growth diminishes as traffic volume approaches the capacity limit. Its implications can be best observed in two extreme cases. When there is no traffic on the motorway, the speed of adjustment is maximum:

$$Y_{it-1} \rightarrow 0 \Rightarrow \tau_{it} \rightarrow 1 \Rightarrow \lambda_{it} \rightarrow \theta \quad (5)$$

In the opposite case, when traffic has reached capacity, the speed of adjustment is zero:

$$Y_{it-1} \rightarrow Y_i^{\max} \Rightarrow \tau_{it} \rightarrow 0 \Rightarrow \lambda_{it} \rightarrow 0 \quad (6)$$

By substituting  $Y_{it}^*$  from Eqs. 1 into 2, we get the first equation:

$$\Delta \ln Y_{it} = \lambda_{it} \cdot (\alpha_i + \beta \cdot \ln X_{it} - \ln Y_{it-1}) + \varepsilon_{it} \quad (7)$$

Next, substituting from  $\lambda_{it}$  for Eq. 4 we get the final equation:

$$\frac{\Delta \ln Y_{it}}{\tau_{it}} = (\theta \cdot \alpha_i + \theta \cdot \beta \cdot \ln X_{it} - \theta \cdot \ln Y_{it-1}) + \frac{\varepsilon_{it}}{\tau_{it}} \quad (8)$$

---

Footnote 3 continued

exogenous to the companies. Presently, toll increases are set according to the Consumer Price Index (CPI) plus an efficiency term.

This is a heteroskedastic model, so we have estimated using weighted least squares. This formulation does not need to be restricted to the partial adjustment model. It can be easily generalised to  $s$  lagged values, as shown in Appendix 1.

We estimate a standard demand equation where variables are expressed in logarithms.<sup>4</sup> The traffic volume in each section is a function of the level of economic activity (measured by Gross Domestic Product, GDP), the toll rate per kilometre, the price of gasoline and a set of dummy variables that capture major changes in the road network.<sup>5</sup> The full set of dummy variables is detailed in Appendix 2. The demand function can be expressed as follows:

$$\frac{\Delta \ln Y_{it}}{\tau_{it}} = (\theta \cdot \alpha_i + \theta \cdot \beta_{1i} \cdot \ln GDP_t + \theta \cdot \beta_{2i} \cdot \ln GP_t + \theta \cdot \beta_{3i} \cdot \ln T_{it} + \theta \cdot \gamma_i \cdot Z_{it} - \theta \cdot \ln Y_{it-1}) + \frac{\varepsilon_{it}}{\tau_{it}} \quad (9)$$

where  $Y_{it}$  is the traffic volume on motorway section  $i$  in period  $t$ ;  $GDP_t$  is the real GDP in period  $t$ ;  $GP_t$  is the gasoline price in period  $t$  deflated by Consumer Price Index, CPI;  $T_{it}$  is the motorway toll in section  $i$  period  $t$  deflated by CPI;  $Z_{it}$  is the dummy variables capturing major changes in the network;  $\alpha_i$  is the individual fixed effects;  $\varepsilon_{it}$  is the error term;  $(\theta, \alpha_i, \beta_{1i}, \beta_{2i}, \beta_{3i}, \gamma_i)$  are the coefficients to be estimated.

The individual fixed effects explain the differences between motorway sections (cross-section units) not captured by the variables included in the model. In our case, they may capture generation and attraction effects that determine the magnitude of traffic in each motorway section.

## The data

To estimate the demand equation, we used a panel data set of 67 motorway sections observed between 1980 and 2008, although not all cross-section units were observed for this temporal span. The total number of observations was 1765. The cross-section observations correspond to the shortest motorway section allowed by the data collection processes, with an average length of 20 km.

The dependent variable is the annual average daily traffic volume in each section. The explanatory variables are: real GDP, gasoline price and toll per km. The last two deflated by CPI. GDP and gasoline price are defined at the national level and take the same value for all sections in the sample.<sup>6</sup> Finally, a set of 30 dummy variables captures the most important changes in the road network. For example, improvements on a parallel free road were captured by a dummy variable that takes value 1 since the opening year. The

<sup>4</sup> We considered the three alternatives most widely used to estimate aggregate demand functions: the linear model, the semi-log model and the log-linear model. According to the Schwarz criterion, based on the log of the likelihood functions from each model, we selected the log-linear specification.

<sup>5</sup> It must be noted that equation (9) could be reparametrised so that the relationship between traffic flow and the explanatory variables can be expressed in levels. Matas and Raymond (2003) provide a justification for this model specification.

<sup>6</sup> In some preliminary estimations we used GDP and gasoline prices at the regional level. The estimated coefficients and the degree of adjustment showed to be almost the same. Therefore, given that for series defined at national level the available time span is much larger, we decided to use GDP and gasoline prices at the national level in order to obtain better forecasting models for the input variables.

**Table 1** Descriptive statistics

	Mean	Maximum	Minimum	Std. dev.
Traffic volume	16807	90033	1689	13523
GDP (millions of €) <sup>a</sup>	733009	1063202	471466	177785
Gasoline price (€ per l) <sup>a</sup>	0.982	1.496	0.832	0.176
Toll (€ per km) <sup>a</sup>	0.126	0.343	0.058	0.050
Maximum capacity	78700	121192	59700	15407

<sup>a</sup> The base year for variables expressed in € is 2006

advantage of working with a panel data set is the high variability observed in the sample. See Table 1.

It is interesting to note that there are substantial differences in traffic volume among the different sections of the motorway network. The daily average traffic flow ranges from 1689 vehicles in the section and year having the lowest volume to 90033 in the section and year with the highest. Furthermore, we found an extensive price range for toll rates. For the whole period, at 2006 prices, the lowest price paid per km was about 0.058 €, whereas the highest was about 0.34 €. The reasons for this wide variation are twofold. Firstly, each motorway has to cover its own construction costs, so the toll rates are higher on those motorways with larger construction costs or lower traffic volume. Secondly, the changes in toll policies during the last two decades have resulted in a wide variation of rates across the country and over time. For instance, on some motorway sections tolls decreased as much as 40% in one year.

The maximum capacity of each motorway section was calculated according to the number of lanes and types of vehicle.

## Model estimation and results

Before estimating the model equation stated in (9), and in order to decide whether to estimate in levels or differences, we analyzed the existence of unit roots and cointegration of the series. The traffic volume and GDP variables were clearly non-stationary. So, both variables could be considered as integrated which means that the expected value and the variance are non constant. The evidence for motorway tolls and gasoline prices was more doubtful. In any case, to justify an estimation using levels for all the variables, it is necessary to guarantee that a cointegration relation exists among them. This means it is possible to find a linear combination of the series that is stationary. In our case, according to the Kao cointegration test for panel data, the null hypothesis of no cointegration was clearly rejected.<sup>7</sup> Therefore, we proceeded to estimate the equation in levels.

As specified in Eq. 9, the estimation of the demand equation would require to estimate 400 coefficients. Given that the number of total observations was 1765, it seemed advisable to introduce some constraints to the coefficients in order to allow for efficiency gains. Based on a previous work by Matas and Raymond (2003), we assumed that the demand

<sup>7</sup> The Kao test is based on the analysis of the residuals. The estimated statistic for the Augmented Dickey-Fuller (ADF) test for the residuals is  $-4.677$ , which clearly rejects the null hypothesis of no cointegration ( $p$ -value almost zero). For a clear presentation of unit roots and cointegration see Hamilton (1994).

elasticity of GDP and gasoline prices were the same across all motorway sections. Nonetheless, we maintained a specific toll coefficient for each motorway section.

Under these assumptions, we estimated Eq. 9 using weighted least squares. The random disturbance of the equation was modelled as a first order autoregressive process ( $\rho$ ) to control for autocorrelation. The coefficients for GDP, gasoline price, and the lagged value of the dependent variable take the expected sign and were estimated with a high degree of precision. In relation to the toll coefficients, a significant variation across motorway sections was observed. A Chi-square test allowed us to clearly reject the null hypothesis of equality of toll coefficients across all sections. However, the difference in the values of the toll coefficients could be explained by certain motorway characteristics: contiguous sections on the same motorway present very similar elasticities; the more inelastic sections are located on corridors with high traffic volumes, and demand is seen to be more elastic where a good alternative free road exists.

The observed results suggested the possibility of re-estimating the model by introducing the hypothesis of equality of toll coefficients across those motorway sections that showed similar coefficients in the initial model. Hence, we proceed by testing equality constraints among the toll coefficients for those motorways with similar coefficients in the original estimation. Based on the results of the Wald test, the motorway sections were classified into 3 groups as follows:

- Low toll elasticity: sections with toll coefficient between 0 and  $-0.2$ .
- Medium toll elasticity: sections with toll coefficient between  $-0.2$  and  $-0.35$
- High toll elasticity: section with toll coefficients larger than  $-0.35$ .

The final estimation results are detailed in Table 2 and the coefficients for the dummy variables in Appendix 2. As can be observed, the toll coefficients are estimated with a high degree of precision. Given that the variables are log-transformed, the estimated coefficients can be interpreted as short term elasticities. Demand is sensitive to toll variations, although in the short term it is inelastic in all three groups.

**Table 2** Summary of the estimation results

Dependent Variable: $\Delta(\text{Ln}(\text{traffic volume}))/\text{tau}$ Estimation method: weighted least squares			
	Coefficient	Std. Error	<i>t</i> -Statistic
Ln (GDP)	0.7538	0.0403	18.72
Ln (gas price)	-0.3802	0.0157	-24.19
Ln (traffic volume(-1))	-0.6059	0.0226	-26.82
Ln (toll_1)	-0.1549	0.0159	-9.72
Ln (toll_2)	-0.3403	0.0193	-17.62
Ln (toll_3)	-0.4879	0.0276	-17.65
Rho	0.7347	0.0218	33.64
Dummy variables	Yes		
Fixed effects	Yes		
R <sup>2</sup>	0.62		
Observations	1668		

**Table 3** Elasticities with respect to GDP

j (years)	Tau			
	0.1	0.5	0.7	1
0	0.075	0.377	0.528	0.754
1	0.146	0.640	0.832	1.051
2	0.213	0.823	1.006	1.168
3	0.275	0.950	1.107	1.214
4	0.334	1.039	1.165	1.232
5	0.389	1.101	1.199	1.239

To provide an additional insight into the accuracy of our model we compared its forecasting capacity to that of a logistic regression model. Using the same explanatory variables, we estimated a logistic regression with a saturation level equal to maximum capacity. According to the mean square error (MSE) for a dynamic forecast over the period 2000–2008, our approach was clearly preferable to the logistic approach.<sup>8</sup>

An interesting property of the proposed functional form is that it makes it possible to avoid the often unrealistic assumption of constant elasticity. As shown in Appendix 3, demand elasticity with respect to an explanatory variable  $X_k$  depends on the value of  $\tau_{it}$ , that is, it depends on the degree of motorway use. For  $\tau_{it} = \tau^0$ , the elasticity with respect to variable  $X_k$  in period J is given by:

$$\varepsilon_J = \beta_k^* \cdot \frac{(1 - \gamma^{*J+1})}{(1 - \gamma^*)} \quad (10)$$

where  $\beta_k^* = \tau^0 \cdot \beta_k$ , being  $\beta_k$  the coefficient associated to  $X_k$ , and  $\gamma^* = (1 - \tau^0 \cdot \theta)$

As an illustration, we compute the demand elasticity with respect to GDP for different values of  $\tau^0$  and for the first 6 years after the change in the exogenous variable. Elasticities are detailed in Table 3. For  $\tau^0 = 1$ , when the level of traffic approaches 0, short-term elasticity is 0.8; after 5 years, the elasticity tends to the long-term value, 1.24. However, as traffic increases and  $\tau^0$  decreases, demand elasticity becomes less sensitive to GDP variations. For  $\tau^0 = 0.1$ , when traffic flow approaches capacity, short-term elasticity is less than 0.1. The elasticity values computed for  $\tau^0 = 0.7$ , which correspond to the average observed value our sample, are in line with those reported in the literature.

Figure 1 displays the elasticity values for  $\tau^0$  ranging from 0.1 to 1.

For the particular case where  $\tau^0 = 1$ , the coefficients can be interpreted as those in the standard partial adjustment model. The short- and long-term elasticities for all the explanatory variables are reported in Table 4.

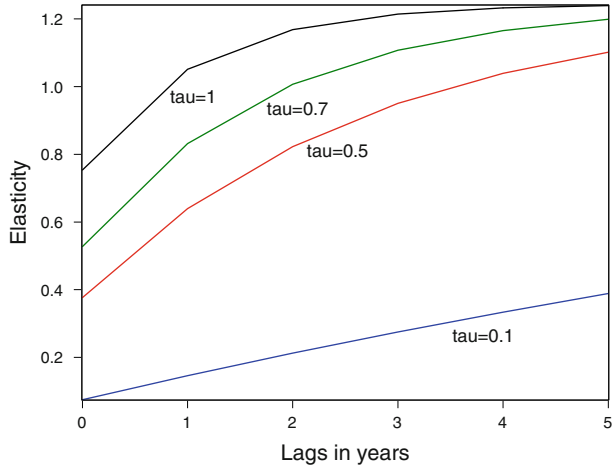
### Forecast results and uncertainty

From the estimated demand model, we proceeded to forecast traffic flow for the 2009–2025 period. The first step was to predict the explanatory variables in the model. GDP and gasoline price are predicted according to a time series model and motorway tolls are

<sup>8</sup> The results showed a value for the MSE of 5,979,872 for the logistic approach and 2,732,930 for our proposal. A “t” test for the equality of both MSE clearly enabled to reject the null hypothesis ( $t = 6.19$ ).



**Fig. 1** Elasticities with respect to GDP for different tau values

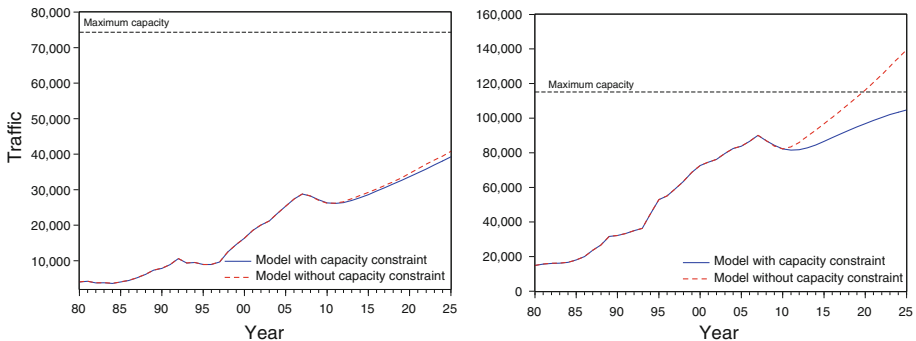


**Table 4** Estimated demand elasticities

	Tau = 1	
	Short term	Long term
GDP	0.754	1.244
Gasoline price	-0.380	-0.628
Toll 1	-0.155	-0.256
Toll 2	-0.340	-0.562
Toll 3	-0.488	-0.805

assumed to remain constant in real terms given that the toll revision formula is linked to CPI. We applied univariate distributions for the exogenous variables given that no correlations were observed among them.

Figure 2 displays the forecasted traffic flow for two representative motorway sections according to both a non-restricted model (standard partial adjustment model) and a capacity restricted model (modified partial adjustment model). In the first one, traffic flow is well below maximum capacity in the year 2025, whereas the second has reached capacity by approximately 2019. As can be observed, the effect of the capacity constraint is



**Fig. 2** Forecasted traffic flow for two motorway sections

almost unnoticeable when traffic volume is below maximum capacity. However, the effect is clear for the second motorway section. The standard partial adjustment predicts an unrealistic level of traffic flow; whereas our suggested functional form forces traffic flow to remain below capacity.

Finally, we proceeded to quantify uncertainty in the traffic forecasts. It is well known that there are three possible sources of error in traffic forecasting. The first one is input uncertainty, due to the fact that the future values of exogenous variables are unknown. The second one is random term uncertainty that accounts for the random disturbance in the demand equation. The third is coefficient uncertainty, due to using parameter estimates instead of true population values. The sum of the last two corresponds to model uncertainty.

To fix ideas, let us consider the following non-linear model:

$$y = \Phi(X, \beta, \varepsilon) \quad (11)$$

in which the dependent variable is, in general, a non-linear function of a set of explanatory variables, of a set of unknown  $\beta$  coefficients and of a random term  $\varepsilon$ . The forecasted values of the dependent variable are obtained by substituting the unknown terms by their respective estimates.

$$\hat{y} = \Phi(\hat{X}, \hat{\beta}, \hat{\varepsilon}) \quad (12)$$

In case we are dealing with a deterministic simulation,  $\hat{\varepsilon}$  is fixed in the expected value of  $\varepsilon$ , that is zero,  $\hat{\beta}$  is the estimated value of  $\beta$ , and  $\hat{X}$  is the assigned value of the explanatory variables.

In a stochastic simulation we assume that each of the elements of Eq. 11 follows a certain distribution. This is:

$$\begin{aligned} X &\sim \text{Dist}(\hat{X}, \Sigma_{\hat{X}}) \\ \beta &\sim \text{Dist}(\hat{\beta}, \Sigma_{\hat{\beta}}) \\ \varepsilon &\sim \text{Dist}(0, \Sigma_{\hat{\varepsilon}}) \end{aligned} \quad (13)$$

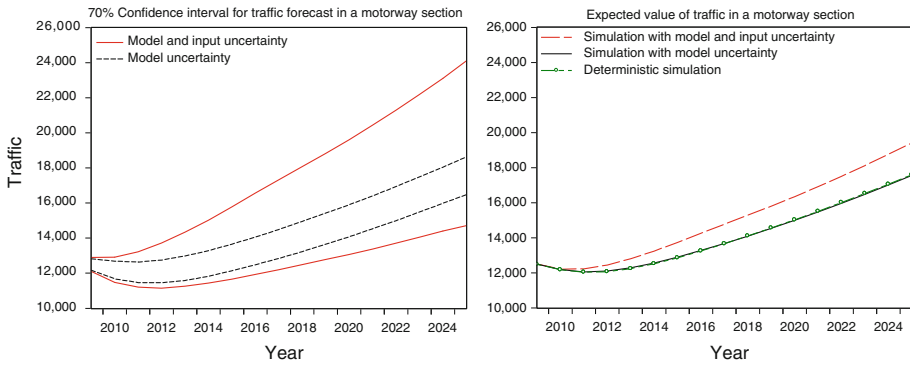
$M$  random realizations of such distributions are generated using a bootstrap methodology. The model is solved for each realization of those distributions. So,  $M$  forecasted values of the dependent variable are obtained. The empirical distribution of the forecasted values enables an expected value to be computed that is the arithmetical average. Using the empirical distribution, for a certain confidence level, it is also possible to compute upper and lower limits. The contribution to total uncertainty derived from the components could be calculated by subtraction. In this study all three types of uncertainty have been obtained through a stochastic simulation process.

To evaluate total forecast uncertainty we consider the distribution of  $\hat{y}$  after generating  $M$  realizations of  $X, \beta, \varepsilon$ .

To evaluate model forecast uncertainty we consider the distribution of  $\hat{y}$  after generating  $M$  realizations of  $\beta, \varepsilon$ ; but holding the values of the explanatory variables  $X$  fixed at  $\hat{X}$ .

Finally, input uncertainty can be computed from the difference between total forecast uncertainty and model forecast uncertainty.<sup>9</sup>

<sup>9</sup> In general, when dealing with non-linear models the additive property is not fulfilled. This means the total forecast uncertainty is not exactly equal to the sum of model forecast uncertainty plus input uncertainty because the existence of interactions or mixed terms uncertainty. Although it would be possible to



**Fig. 3** Confidence intervals and expected traffic flow for one motorway section

Because the model is non-linear it should be noted that the empirical average of the stochastic simulations, in general, will not coincide with the deterministic simulation. Therefore, in non-linear models the deterministic simulation will offer a biased forecast.

In this study the model has been solved repeatedly for 1000 random draws of various components by using bootstrapping method.

To illustrate the impact of uncertainty, we computed the 70% confidence interval for the traffic forecast of one of the motorway sections. As can be observed in Fig. 3, model uncertainty (dashed line) is relatively low and almost constant over time. However, once input uncertainty is added (solid line) the confidence interval widens and clearly increases over time. The second part of Fig. 3 shows the expected value of traffic for a deterministic forecast (dotted line), model uncertainty (solid line) and total uncertainty (dashed line). It can clearly be observed that the deterministic simulation will underpredict the average level of traffic flow.<sup>10</sup>

As previously mentioned and shown in Eq. 11, the model is non-linear and stochastic. Under these conditions, in general, the deterministic solution of a stochastic model will offer a biased estimate of the expected traffic value. Nonetheless, the expected value of the traffic forecast can be approximated by using the average of a set of stochastic simulations. Applying this approach to all the motorway sections in the sample, we found that the stochastic forecast for the year 2025 was on average 8.8% higher than the deterministic forecast.

Table 5 offers an order of magnitude of uncertainty for the same motorway section featured in Fig. 3. The coefficient of variation for total uncertainty ranges from 0.03 in the first forecasted year to 0.24 in the last. In the first few years, uncertainty is low and mainly explained by model uncertainty. However, as time goes by, total uncertainty increases due to lower precision in predicting the unknown values of exogenous variables.

Footnote 9 continued

differentiate between the three components of uncertainty, we have preferred to add the mixed terms to input uncertainty. The reason for proceeding this way is that mixed terms are relatively unimportant compared with the other two components. Therefore, we have computed input uncertainty as total uncertainty minus model uncertainty.

<sup>10</sup> The deterministic solution of a stochastic non-linear model will be biased. When variables are expressed in logarithms, as in our case, the bias will be negative. However, with other functional forms the bias can have a different sign.

**Table 5** Coefficient of variation for total uncertainty and % explained by model and input

	CV	Model (%)	Input (%)
2009	0.032	80.3	19.7
2010	0.059	69.8	30.2
2011	0.082	59.3	40.7
2012	0.103	51.7	48.3
2013	0.121	47.2	52.8
2014	0.136	42.7	57.3
2015	0.150	38.9	61.1
2016	0.163	36.4	63.6
2017	0.174	34.8	65.2
2018	0.183	33.3	66.7
2019	0.191	31.8	68.2
2020	0.200	30.4	69.6
2021	0.209	29.1	70.9
2022	0.217	28.2	71.8
2023	0.224	27.1	72.9
2024	0.232	26.1	73.9
2025	0.242	25.3	74.7

### Uncertainty effects on forecasting forgone revenue

An issue on the political agenda of the Spanish government is to remove tolls on certain motorways before the concession expires. In these cases, the government has had to compensate the private motorway concessionaire for the revenue forgone up to the end of the concession period. We selected one motorway section in the sample in order to compute the effect of uncertainty on the revenue to be forgone. The selected section was 20 kilometres in length with an average traffic value of around 12800 vehicles per day. We assumed that the concession period would expire in 2025.

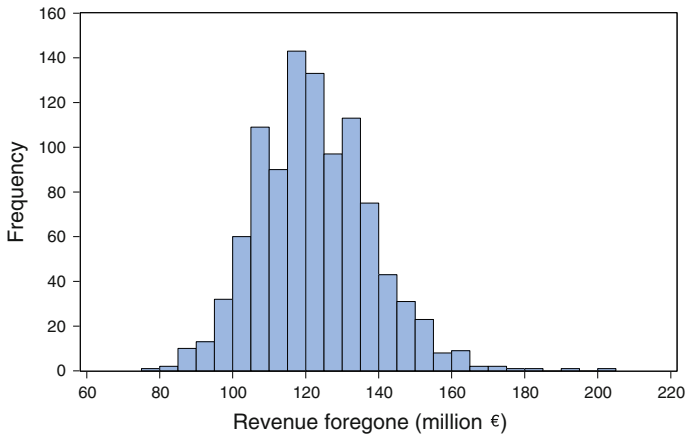
The annual revenue was obtained by multiplying the predicted traffic by the average toll paid by 365 days a year.<sup>11</sup> This value is computed for each forecasted year from 2009 to 2025 and for each of the 1000 random draws. Next, we worked out the results by calculating the Net Present Value (NPV) of the revenue to be forgone along these 17 years at a discounting rate of 5%.

Finally, we analysed the empirical distribution of the NPV, which enabled us to calculate the mean and the confidence intervals for different significance levels. For the selected motorway section, the expected NPV of revenue is 123 million €. The minimum and maximum values for the confidence interval at 70% significance are 107 million € and 138 million €; when we compute the interval at 95% the figures are 94 million € and 155 million €. In the first case, the difference between the two extremes is 29%, whereas in the second it rises to 65%.

Figure 4 presents the empirical distribution of the NPV.

Quantifying uncertainty provides evidence that using point estimates to assess investments or public policies can lead to errors in the decision-making process. In this example,

<sup>11</sup> To obtain the compensation to be paid to the concessionaire, we should deduct from the revenue to be forgone any taxes or other costs related to toll operation.



**Fig. 4** Distribution of the NPV of revenues foregone

the negotiation process between government and concessionaire should include the probabilities associated with the different forecasted revenue values.

## Conclusions

This article contributes to the literature on transport demand forecasting in three different ways: The proposal of a new methodology to account for capacity constraints in long term forecasting, the analysis of the role played by the different components of uncertainty, and the importance of using stochastic simulation techniques to avoid forecasting bias in non-linear models.

Firstly, the proposed functional allows handling existing restrictions on the capacity in those cases where it is not possible to jointly estimate the demand and supply side of the model. Our approach makes it possible to account for capacity constraints in long term forecasting without imposing an arbitrary functional form. This is achieved by specifying a dynamic model in which the speed of adjustment is related to the ratio between the actual traffic flow and the maximum capacity of the motorway. Furthermore, with this functional form, demand elasticity is not constant but depends on the degree of motorway use. As traffic increases, and approaches maximum capacity, demand becomes less sensitive to changes in the explanatory variables.

With respect to uncertainty, this article outlines the importance of developing stochastic simulations based on bootstrapping methodologies in order to obtain confidence intervals for the forecast. The results confirm that in the first few years model uncertainty explains most of the range of variation for the forecast traffic flow. However, as time goes by, whereas model uncertainty remains almost constant, input uncertainty steadily increases so that at the end of the forecasting period the last one accounts for almost 75% of total variability. Based on previous experiences and on the results of this article, it can be concluded that more effort must be made to improve model specification and to implement the necessary mechanisms to avoid bias in forecasting. Nevertheless, input uncertainty has proved to be the main factor for explaining uncertainty in the long run. Consequently, our study shows that forecasting explanatory variables deserves special attention. So it would be advisable to avoid introducing explanatory variables difficult to predict, although these variables might increase the level of adjustment of the model.

Finally, for non-linear models this article calls attention to the inadequacy of the deterministic simulation to forecast future traffic volumes. When dealing with non-linear models, the expected future traffic value can be approximated by averaging the different realizations of the variable using stochastic simulations. As an illustration, this article shows that the deterministic simulation at the end of the forecasting period underpredicts expected traffic flow across all motorway sections in the sample by on average 9% with a maximum difference of 12%.

**Acknowledgments** This work has benefit from a research grant from the Spanish Ministerio de Fomento-CEDEX (PT2007-001-IAPP) as well as from the project ECO2009-12234. We would like also to thank the researchers of the CEDEX project and two anonymous reviewers for their valuable comments.

**Appendix 1. Generalisation to  $s$  lags**

The dynamic partial adjustment model can be easily generalised to account for  $s$  lags as follows:

The static equilibrium equation takes the standard form:

$$\ln Y_{it}^* = \alpha + \beta \cdot \ln X_{it} \tag{14}$$

whereas for the dynamic part we assume that the adjustment process is a weighted function of  $s$  lags:

$$\begin{aligned} \Delta \ln Y_{it} = & \lambda_{it} \cdot \left[ w_1 \cdot (\ln Y_{it}^* - \ln Y_{it-1}) + w_2 \cdot (\ln Y_{it}^* - \ln Y_{it-2}) \cdots \right. \\ & \left. + w_s \cdot (\ln Y_{it}^* - \ln Y_{it-s}) \right] + \varepsilon_{it} \tag{15} \\ \sum_{i=1}^s w_i = & 1 \end{aligned}$$

That is, the correction of the disequilibrium between the actual value and the optimal or desired value of the dependent variable at period  $t$  depends on the disequilibrium in the previous periods. Substituting the expression for  $Y^*$  defined in (14) into Eq. 15, we obtain:

$$\begin{aligned} \Delta Y_{it} = & \lambda_{it} \cdot [w_1 \cdot (\alpha + \beta \cdot X_{it} - Y_{it-1}) + w_2 \cdot (\alpha + \beta \cdot X_{it} - Y_{it-2}) \cdots + w_s \cdot (\alpha + \beta \cdot X_{it} - Y_{it-s})] + \varepsilon_{it} \\ = & \lambda_{it} \cdot [\alpha + \beta \cdot X_{it} - w_1 \cdot Y_{it-1} - w_2 \cdot Y_{it-2} - w_3 \cdot Y_{it-3} - \cdots - w_{s-1} \cdot Y_{it-s+1} - Y_{it-s}] + \varepsilon_{it} \\ = & \lambda_{it} \cdot \alpha + \lambda_{it} \cdot \beta \cdot X_{it} - \lambda_{it} \cdot w_1 \cdot Y_{it-1} - \lambda_{it} \cdot w_2 \cdot Y_{it-2} - \cdots - \lambda_{it} \cdot w_{s-1} \cdot Y_{it-s+1} - \lambda_{it} \cdot Y_{it-s} + \varepsilon_{it} \end{aligned}$$

Given that

$$\lambda_{it} = \theta \cdot \left( \frac{Y_i^{\max} - Y_{it-1}}{Y_i^{\max}} \right) = \theta \cdot \tau_{it}$$

and substituting:

$$\begin{aligned} \Delta \ln Y_{it} = & \theta \cdot \tau_{it} \cdot \alpha + \theta \cdot \tau_{it} \cdot \beta \cdot \ln X_{it} - \theta \cdot \tau_{it} \cdot w_1 \cdot \ln Y_{it-1} - \theta \cdot \tau_{it} \cdot w_2 \cdot \ln Y_{it-2} \\ & - \cdots - \theta \cdot \tau_{it} \cdot w_{s-1} \cdot \ln Y_{it-s+1} - \theta \cdot \tau_{it} \cdot \ln Y_{it-s} + \varepsilon_{it} \end{aligned}$$

That is:

$$\begin{aligned} \frac{\Delta \ln Y_{it}}{\tau_{it}} = & \theta \cdot \alpha + \theta \cdot \beta \cdot \ln X_{it} - \theta \cdot w_1 \cdot \ln Y_{it-1} - \theta \cdot w_2 \cdot \ln Y_{it-2} - \cdots \\ & - \theta \cdot w_{s-1} \cdot \ln Y_{it-s+1} - \theta \cdot \ln Y_{it-s} + \frac{\varepsilon_{it}}{\tau_{it}} \end{aligned}$$

## Appendix 2

See Table 6

**Table 6** Definition of dummy variables and their estimated coefficients

Dummy variables	Period	Comment	Coefficient
$Z_1$	2006–2008	It reflects the positive impact on traffic on AP-1 motorway after removing the toll on the connecting city ring-road.	0.1039
$Z_2$ – $Z_4$	2004–2008	They reflect the negative impact on traffic on 3 AP-2 motorway sections as a consequence of the extension of an alternative toll free road.	–0.1989 –0.2187 –0.2142
$Z_5$ – $Z_6$	1992	They account for the positive impact on the 2 AP-4 motorway sections, derived from the Seville World Exhibition in 1992.	0.1828 0.1934
$Z_7$	2005–2008	It reflects the positive impact on traffic on the 2 AP-4 motorway sections after removing the toll in the final section of the motorway.	0.1371
$Z_8$ – $Z_9$	2003–2008	They reflect the negative impact on traffic on the 2 AP-68 motorway sections, derived from quality improvements on the alternative toll free road.	–0.0337 –0.0700
$Z_{10}$ – $Z_{13}$	1994–2008	They reflect the negative impact on traffic on 4 AP-7 motorway sections, derived from the extension of an alternative tollway.	–0.0735 –0.0990 –0.1048 –0.0660
$Z_{14}$	2000–2008	It reflects the positive impact on traffic on 1 AP-7 motorway section after removing the toll for the connecting city's ring road.	0.0653
$Z_{15}$	1999–2000	It reflects the positive impact on traffic on 1 AP-7 motorway section, due to the toll exemption because of construction works on the motorway.	0.1347
$Z_{16}$ – $Z_{17}$	1998–2008	They reflect the negative impact on traffic on 2 AP-7 motorway sections, derived from the opening of an alternative toll motorway.	–0.1453 –0.1203
$Z_{18}$	1998–2008	It reflects the positive impact on traffic on 1 AP-7 section after removing the toll in the city ring-road.	0.1082
$Z_{19}$ – $Z_{20}$	2006	They reflect the positive impact on traffic on 2 AP-7 motorway sections, due to the road works on the alternative toll free road	0.0636 0.1468
$Z_{21}$ – $Z_{22}$	2007–2008	They reflect the negative impact on traffic on 2 AP-7 motorway sections as a consequence of the extension of an alternative toll motorway.	–0.1540 –0.1578

**Table 6** continued

Dummy variables	Period	Comment	Coefficient
Z <sub>23</sub>	2004–2008	They reflect the negative impact on traffic on 1 AP-7 motorway section, derived from the opening of an alternative toll free motorway.	−0.2321
Z <sub>24</sub> –Z <sub>26</sub>	1993–2008	They reflect the negative impact on traffic on 3 AP-7 motorway sections, derived from the opening of an alternative toll free motorway.	−0.0821 −0.0533 −0.0555
Z <sub>27</sub>	2004–2007	It reflects the negative impact on traffic on 1 A-8 motorway section, derived from the opening of a connection with the alternative toll free motorway.	−0.2125
Z <sub>28</sub> –Z <sub>30</sub>	2003–2007	They reflect the negative impact on traffic on 3 A-8 motorway sections as a consequence of the extension of an alternative toll free motorway.	−0.0915 −0.1212 −0.1274

**Appendix 3. Demand elasticity with a capacity constraint**

This appendix shows how demand elasticity depends on the level of traffic when a constraint on infrastructure capacity is in force.

Starting from Eq. 8 in the main text:

$$\frac{\Delta \ln Y_{it}}{\tau_{it}} = (\theta \cdot \alpha_i + \theta \cdot \beta \cdot \ln X_{it} - \theta \cdot \ln Y_{it-1}) + \frac{\varepsilon_{it}}{\tau_{it}}$$

Defining betas and taking expected values, we get:

$$\frac{\Delta \ln Y_{it}}{\tau_{it}} = \beta_0 + \beta_1 \cdot \ln X_{it} - \theta \cdot \ln Y_{it-1}$$

where  $\beta_0 = \theta \cdot \alpha_i$  and  $\beta_1 = \theta \cdot \beta$

It can be rewritten as:

$$\ln Y_{it} = \tau_{it} \cdot \beta_0 + \tau_{it} \cdot \beta_1 \cdot \ln X_{it} + (1 - \tau_{it} \cdot \theta) \cdot \ln Y_{it-1}$$

$$\ln Y_{it} = \beta_{0it}^* + \beta_{1it}^* \cdot \ln X_{it} + \gamma_{it}^* \cdot \ln Y_{it-1}$$

where

$$\beta_{kit}^* = \tau_{it} \cdot \beta_k; \gamma_{it}^* = (1 - \tau_{it} \cdot \theta)$$

For a fixed level of traffic flow,  $\tau_{it} = \tau^0$ , we have:

$$\ln Y_{it} = \beta_0^* + \beta_1^* \cdot \ln X_{it} + \gamma^* \cdot \ln Y_{it-1}$$

$$\beta_k^* = \tau^0 \cdot \beta_k; \gamma^* = (1 - \tau^0 \cdot \theta)$$

Taking into account the dynamic structure of the model and following a recursive process of substitution (this is, substituting  $\ln Y_{it-1} = \beta_0^* + \beta_1^* \cdot \ln X_{it-1} + \gamma^* \cdot \ln Y_{it-2}$ ), the elasticity in period J will be:



$$\varepsilon_J = \beta_k^* \cdot \frac{(1 - \gamma^{*J+1})}{(1 - \gamma^*)}$$

## References

- Bain, R.: Error and optimism bias in toll road traffic forecasts. *Transportation* **36**, 469–482 (2009)
- de Jong, G., Daly, A., Pieters, M., Miller, S., Plasmeijer, R., Hofman, F.: Uncertainty in traffic forecasts: literature review and new results for The Netherlands. *Transportation* **34**, 375–395 (2007)
- Flyvbjerg, B.: Curbing optimism bias and strategic misrepresentation in planning: reference class forecasting in practice. *Eur. Plan. Stud.* **16**, 3–21 (2008)
- Flyvbjerg, B., Skamris, M., Buhl, S.: Inaccuracy in traffic forecasts. *Transp. Rev.* **26**, 1–24 (2006)
- Hamilton, J.D.: *Time Series Analysis*. Princeton University Press, Princeton (1994)
- Matas, A., Raymond, J.L.: Demand elasticity on tolled motorways. *J. Transp. Stat.* **6**, 91–108 (2003)
- Riddington, G.: Long range air traffic forecasts for the UK: a critique. *J. Transp. Econ. Policy* **40**(2), 297–314 (2006)
- Vassallo, J., Baeza, M. Why traffic forecasts in PPP contracts are often overestimated? Research Paper, Final Draft, EIB University Research Sponsorship Programme, European Investment Bank, Luxembourg (2007)

## Author Biographies

**Anna Matas** is professor of Applied Economics at the Universitat Autònoma de Barcelona. Her main research interests are in the areas of transport economics, regulation and impact of infrastructure on location. She has several articles published in academic journals in the field of transport economics. She has also coauthored some books on topics related to the Spanish Economy.

**Josep-Lluís Raymond** is professor of Econometrics at the Universitat Autònoma de Barcelona. His main fields of interest are econometrics, human capital, transport and infrastructures. He is author of several books and articles published in academic journals. He has also done consultancy work.

**Adriana Ruiz** is a PhD candidate at the Universitat Autònoma de Barcelona. Her doctoral dissertation focuses on analyzing several economic aspects of transport infrastructure, such as the impact of toll roads on consumer welfare, and the effects of transport infrastructure investment on the location of economic activities in Spain.