Traffic forecasts under uncertainty and capacity constraints

Anna Matas
Dpt. Economia Aplicada. Universitat Autònoma de Barcelona
Josep-Lluis Raymond
Dpt. Economia i Història Econòmica. Universitat Autònoma de Barcelona
Adriana Ruiz
Dpt. Economia Aplicada. Universitat Autònoma de Barcelona

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 paper 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 paper 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. 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

February 2010

1. **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.

This paper 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 includes both the variance of the traffic forecast related to the stochastic character of the model (model uncertainty) and the uncertainty that underlines the future values of the exogenous variables (input uncertainty). Furthermore, we apply this methodology to a specific public policy consisting of suppressing the toll on a certain motorway section 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 standard variables. 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 paper contributes to this issue by proposing a new functional form for the demand equation that accounts for the fact that the rate of traffic flow growth diminishes as it approaches full capacity. Specifically, as detailed in section 3, we suggest a modified partial adjustment model with variable adjustment speed. In our case, this proposal outweighs the traditional logistic functional form with a saturation level equal to maximum capacity, given that we avoid the assumption that traffic follows and S-shaped growth curve.

2. **Literature review of uncertainty in traffic forecasting**

Several recent studies confirm the inaccuracy of traffic predictions. Among them, the extensive work by Flyvbjerg, Skamris Holm and Buhl (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 ±20%[1]. 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 toll road traffic forecasting performance 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%.

[1] 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 was overestimated by approximately 35% during the first three years of operation. They conclude that there is a substantial optimism bias in the ramp-up period for toll motorway concessions in Spain.

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 actual likelihood of this 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 link flows between 18% and 33% of the mean. 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 paper intends to contribute to this issue with new findings.

3. **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 traffic rate of growth diminishes as it approaches maximum capacity.

Ideally, congestion costs and capacity constraints should be considered by using an integrated demand-supply system. However, this is frequently unfeasible. As an 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 equilibrium value of traffic Y^* as a function of a set of variables X:

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

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

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

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, τ, as a function of the traffic flow related to the maximum capacity of the infrastructure, Y^0:

$$\tau_{it} = \frac{Y_i^0 - Y_{it-1}}{Y_i^0} \tag{3}$$

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

$$\lambda_{it} = \theta\left(\frac{Y_i^0 - Y_{it-1}}{Y_i^0}\right) = \theta \cdot \tau_{it} \tag{4}$$

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

$$Y_{it-1} = 0 \rightarrow \tau_{it} = 1 \rightarrow \lambda_{it} = \theta \tag{5}$$

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

$$Y_{it-1} = Y_i^0 \rightarrow \tau_{it} = 0 \rightarrow \lambda_{it} = 0 \tag{6}$$

By substituting ln*Y** from equation (1) into (2), we get the first equation:

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

Next, substituting λ_{it} for its value we get the final equation:

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

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*” lags by assuming that the adjustment process is a weighted function of “*s*” lags, as shown in Appendix 1.

We estimate a standard demand equation where variables are expressed in logs[2]. 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[3]. The demand function can be expressed as follows:

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

Where Y_{it} = traffic volume on motorway section *i* in period *t*

GDP_t = real GDP in period *t*

GP_t = gasoline price in period *t* deflated by Consumer Price Index, CPI)

T_{it} = motorway toll in section *i* period *t* deflated by CPI

Z_{it} = dummy variables capturing major changes in the network

α_i = individual fixed effects

ε_{it} = error term

The individual fixed effects explain the differences between cross-section observations 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.

4. **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 kilometres.

[2] 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 criterion, based on the log of the likelihood functions from each model, we selected the log-linear specification.

[3] Matas and Raymond (2003) provide a justification for this model specification.

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[4]. Finally, a set of 31 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 main descriptive statistics for the variables are outlined in Table 1.

Table 1. Descriptive statistics

	Mean	Maximum	Minimum	Std. Dev.
Traffic volume	16807	90033	1689	13523
GDP (millions of €)[1]	733009	1063202	471466	177785
Gasoline price (€ per liter)[1]	0.982	1.496	0.832	0.176
Toll (€ per km)[1]	0.126	0.343	0.058	0.050
Maximum capacity	78700	121192	59700	15407

[1] The base year for variables expressed in euros is 2006.

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.

[4] 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.

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

5. 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 the cointegration of the series. The traffic volume and GDP variables were clearly non-stationary. The evidence for motorway tolls and gasoline prices was doubtful. In any case, to justify an estimation using levels of those variables, it was necessary to guarantee that cointegration relations existed among them. Using the Kao test for panel data and based on the analysis of residuals, the null hypothesis of no cointegration was clearly rejected. Given that this hypothesis was rejected, we proceeded to estimate the equation in levels.

Table 2. Kao Residual Cointegration Test

Series: LOG(traffic_?) LOG(toll_?) LOG(GDP) LOG(gas price)
Sample: 1980-2008
Included observations: 29
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
User-specified lag length: 1
Newey-West automatic bandwidth selection and Bartlett kernel

	t-Statistic	Prob.
ADF	-4.676612	0.0000
Residual variance	0.001894	
HAC variance	0.003072	

As specified in equation (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 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 equation (9) using weighted least squares. An AR(1) term was included to control autocorrelation of the error term. 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 characteristics; 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. We re-estimated the initial model introducing equality constraints among coefficients not rejected by the data. The constraints were introduced by classifying the motorway sections into 3 groups according to the toll coefficient estimated in the general model:

- 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 3. As can be observed, the toll coefficients are estimated with a high degree of precision. Demand is sensitive to toll variations, although in the short term it is inelastic in all three groups.

To provide an additional insight on 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. We compute the mean square error (MSE) for a dynamic forecast over the period 2000-2008. The results showed a value 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).

Table 3. Estimation results

Dependent Variable: D(Ln(traffic))/tau
Estimation method: weighted least squares

	Coefficient	**Std. Error**	**t-Statistic**
ln(GDP)	0.753772	0.040268	18.72
ln(Gas price)	-0.380198	0.015718	-24.19
ln(traffic(-1))	-0.605873	0.022587	-26.82
ln(toll_1)	-0.154903	0.015938	-9.72
ln(toll_2)	-0.340256	0.019311	-17.62
ln(toll_3)	-0.487923	0.027644	-17.65
AR(1)	0.734674	0.021841	33.64
Dummy variables	yes		
Fixed effects	yes		
R^2	0.62		
Observations	1668		

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 2, demand elasticity depends on the value of τ, that is, it depends on the degree of motorway use. For a given level of τ, elasticity in period J is defined as:

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

Where $\beta_k^* = \tau \cdot \beta_k$ and $\gamma^* = (1 - \tau \cdot \theta)$

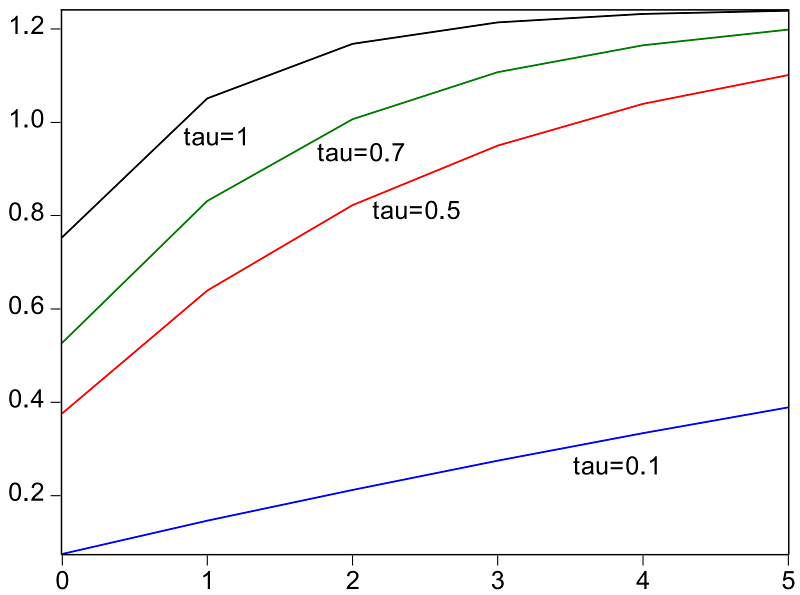
As an illustration, we compute the demand elasticity with respect to GDP for different values of τ and for the first 6 years after the change in the exogenous variable. Elasticities are detailed in Table 4. For τ = 1, when the level of traffic approaches 0, short-term elasticity is 0.8; after 5 years, the elasticity tends to the long-term value, 1.24. However, as traffic increases and τ decreases, demand elasticity becomes less sensitive to GDP variations. For τ = 0.1, when traffic flow approaches capacity, short-term elasticity is less than 0.1. The elasticity values computed for τ = 0.7, which correspond to the average observed value in our sample, are in line with those reported in the literature.

Table 4. Elasticities with respect to GDP

	tau			
j (years)	**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

Figure 1 displays the elasticity values for τ ranging from 0.1 to 1.

Figure1. Elasticities with respect to GDP for different tau values



For the particular case where τ = 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 5.

Table 5. Estimated demand elasticities

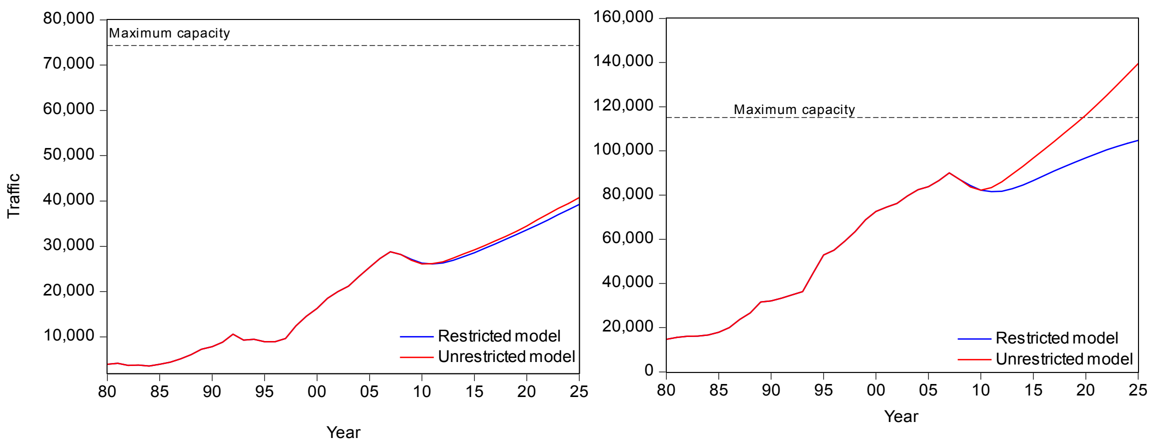
	tau = 1	
	Short Term	**Long Term**
GDP	0.754	1.244
Gasoline price	-0.380	-0.628
Toll 1	-0.155	-0.256
Toll 2	-0.340	-0.562
Toll 3	-0.488	-0.805

6. 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 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 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.

Figure 2. Forecasted traffic flow for two motorway sections



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 specification errors in the demand equation. The third is coefficient uncertainty, due to using parameter estimates instead of true values. The sum of the last two corresponds to model uncertainty.

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

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

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

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

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

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

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

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, is also possible to compute upper and lower limits. The contribution to total uncertainty derived from the components could be calculated by difference. 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, β, ε.

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

Finally, input uncertainty can be computed from the difference between total forecast uncertainty and model forecast uncertainty.

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.

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 Figure 3, model uncertainty (black line) is relatively low and almost constant over time. However, once input uncertainty is added (red line) the confidence interval widens and clearly increases over time. The second part of Figure 3 shows the expected value of traffic for a deterministic forecast (green line), model uncertainty (black line) and total uncertainty (red line). It can clearly be observed that the deterministic simulation will underpredict the average level of traffic flow.

As previously mentioned and shown in equation (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.

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

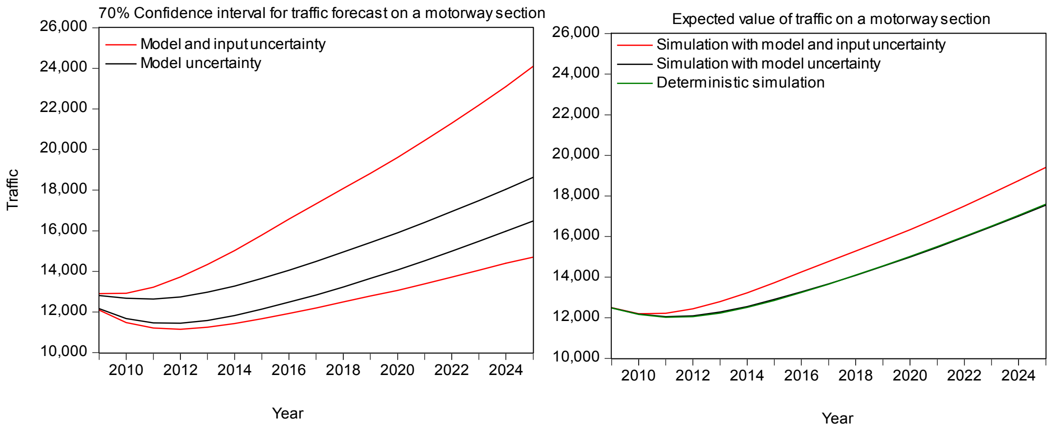


Table 6 offers an order of magnitude of uncertainty for the same motorway section featured in Figure 3. The coefficient of variation for total uncertainty ranges from 0.03 in the first forecasted year to 0.24 in the last. In the first few years, uncertainty is low and mainly explained by model uncertainty. However, as time goes by, total uncertainty increases due to lower precision in predicting the unknown values of exogenous variables.

Table 6. 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%

7. Uncertainty effects on forecasting forgone revenue

In recent year, tolls on certain motorway sections have been removed 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.

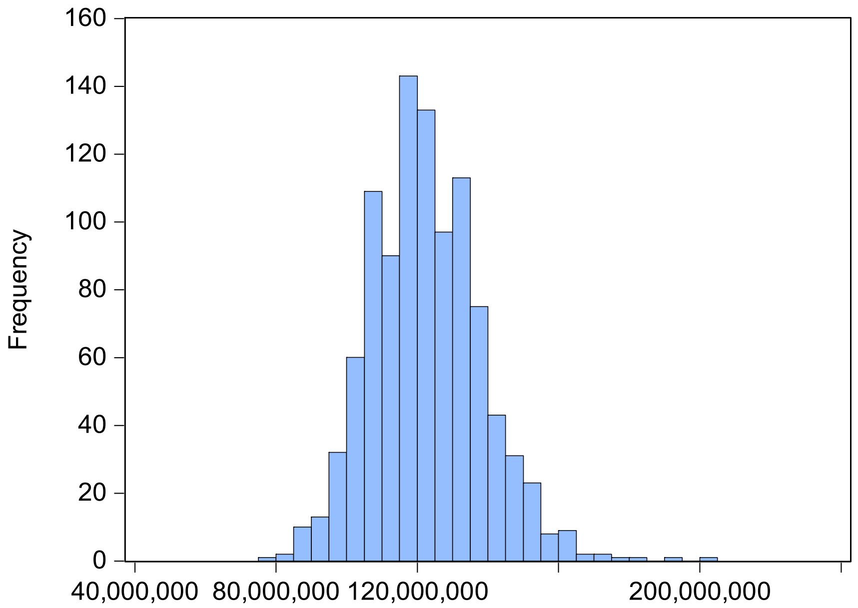
The annual revenue was obtained by multiplying the predicted traffic by the average toll paid by 365 days a year[5]. This value is computed for each forecasted year from 2009 to 2025 and for each of the 1000 random draws. Next, we worked out the results by calculating the Net Present Value (NPV) of the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

[5] 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.

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 122,897,992 euros. The minimum and maximum values for the confidence interval at 70% significance are 106,595,480 euros and 137,934,400 euros; when we compute the interval at 95% the figures are 93,952,016 euros and 154,969,120 euros. 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.

Figure 4. Distribution of the NPV of revenues forgone



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

8. Conclusions

This article proposes a new methodology for incorporating capacity constraints into long-term traffic predictions. Doubtless, a complete model that simultaneously considers traffic demand and traffic supply would be the best solution. However, frequently, not enough data is available to construct the supply side of the model. In such cases, our approach makes it possible to carry out long-term forecasts that verify the capacity constraints of the motorway without having to introduce 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.

This paper also outlines the importance of developing stochastic simulations based on bootstrapping methodologies in order to obtain confidence intervals for the forecast. In a traffic demand model, the uncertainty of the forecast depends on the uncertainty of the estimated coefficients and on the uncertainty coming from the random disturbance term of the model. It also depends on uncertainty about the future values of the explanatory variables. In general, the latter type of uncertainty is the most important. By integrating the three types of uncertainty through the stochastic simulation process, it is possible to obtain confidence intervals for future traffic volumes. In the case of toll motorways, those traffic intervals enable computing confidence intervals for the present value of the future toll revenue.

Finally, for non-linear models this paper 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 paper 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%.

9. References

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

de Jong, G., A. Daly, M. Pieters, S. Miller, R. Plasmeijer, and F. Hofman (2007): “Uncertainty in traffic forecasts: literature review and new results for The Netherlands”, *Transportation,* 34 , 375–395.

Flyvbjerg, B. (2008): “Curbing Optimism Bias and Strategic Misrepresentation in Planning: Reference Class Forecasting in Practice”, *European Planning Studies*, 16, 3-21.

Flyvbjerg, B., M. Skamris and S.Buhl (2006): “Inaccuracy in Traffic Forecasts”, *Transport Reviews*, 26, 1 - 24.

Matas, A. and J.L. Raymond (2003): “Demand Elasticity on Tolled Motorways”, *Journal of Transportation and Statistics*, 6 , 91-108.

Vassallo, J. and M. Baeza (2007): “Why Traffic Forecasts in PPP Contracts Are Often Overestimated?” *Research Paper, Final Draft, EIB University Research Sponsorship Programme,* European Investment Bank, Luxembourg.

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:

$$\ln Y_{it}^{*} = \alpha + \beta \cdot \ln X_{it} \tag{1.1}$$

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

$$\Delta \ln Y_{it} = \lambda_{it} \cdot \left[w_1 \cdot (\ln Y_{it}^{*} - \ln Y_{it-1}) + w_2 \cdot (\ln Y_{it}^{*} - \ln Y_{it-2}) \cdots\cdots + w_s \cdot (\ln Y_{it}^{*} - \ln Y_{it-s}) \right] + \varepsilon_{it}$$

$$\sum_{i=1}^{s} w_i = 1 \tag{1.2}$$

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 (1.1) into equation (1.2), we obtain:

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

Given that

$$\lambda_{it} = \theta \cdot \left(\frac{Y_i^0 - Y_{it-1}}{Y_i^0} \right) = \theta \cdot \tau_{it}$$

and substituting:

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

That is:

$$\frac{\Delta \ln Y_{it}}{\tau_{it}} = \theta \cdot \alpha_i + \theta \cdot \beta \cdot \ln X_{it} - \theta \cdot w_1 \cdot \ln Y_{it-1} - \theta \cdot w_2 \cdot \ln Y_{it-2} - \cdots\cdots - \theta \cdot w_{s-1} \cdot \ln Y_{it-s+1} - \theta \cdot \ln Y_{it-s} + \frac{\varepsilon_{it}}{\tau_{it}^{p}}$$

Appendix 2. 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.

For simplicity, we assume that the modified partial adjustment model is:

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

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

$$\beta^*_{kit} = \tau_{it} \cdot \beta_k \qquad \gamma^*_{it} = (1 - \tau_{it} \cdot \theta)$$

For a given level of traffic flow, τ, we have:

$$\ln Y_{it} = \beta^*_0 + \beta^*_1 \cdot \ln X_{it} + \gamma^* \cdot \ln Y_{it-1}$$

$$\beta^*_k = \tau \cdot \beta_k \qquad \gamma^* = (1 - \tau \cdot \theta)$$

Taking into account the dynamic structure of the model and following a recursive process of substitution, the elasticity in period J will be:

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