Traffic forecast: capacity constraints and uncertainty

A. Matas, J.L.Raymond and A. Ruiz
Universitat Autònoma de Barcelona

4th Kuhmo-Nectar Conference
Copenhague, July 2009

1. Introduction

- Aggregate transport models are regularly used for long term traffic forecasts
- Objective of the paper is to address two problems related to forecasting:
  - Capacity constraints
  - Uncertainty
- Tolled Motorway Traffic in Spain

2. Empirical evidence related to uncertainty

1. Recent studies confirm the high level of inaccuracy in traffic demand forecasts

Flyvbjerg et al. (2005), (2006), (2008).

Bain and Wilkins (2002), Bain (2004); Bain and Polakovic (2005).

2. De Jong et al. (2007) provide an extensive review of studies that quantify uncertainty in the traffic forecasts. The paper also offers a methodology for quantifying uncertainty and an application in The Netherlands.

3. The data

- 69 motorway sections observed between 1980 and 2008.
- 1846 observations.
- Dependent variable: average daily traffic volume in each section.
- Explanatory variables: GDP, petrol price, toll and dummy variables accounting for changes in the network.

3. The data

Descriptive statistics

	Mean	Maximum	Minimum	Std. Dev.
Traffic volume	17,453	90,033	1,689	14,481
GDP (millions of €)	732,923	1,063,202	471,466	177,809
Gasoline price (€ per l.)	0.982	1.496	0.832	0.176
Toll (€ per km)	0.1267	0.3434	0.0579	0.0500

4. The model

The demand equation can be expressed as follows:

$$Y_{it} = f(GDP_t, GP_t, T_{it}, Z_{it}, \alpha_i, \varepsilon_{it})$$

GPD_t: Real GDP in period t

GP_t: Gasoline price in period t

T_{it}: Motorway toll in section i period t

Z_{it} : Dummy variables accounting for major changes in the network

α_i: Fixed effects

ε_{it}: Error term

4. The model

Null Hypothesis of no cointegration is rejected

Kao Residual Cointegration Test
Series: LOG(I_?) LOG(P_?) LOG(PIBES) LOG(GAS)
Sample: 1980 2008
Included observations: 29
Null Hypothesis: No cointegration
Trend assumption: No deterministic trend
Lag selection: fixed at 1
Newey-West bandwidth selection using Bartlett kernel

	t-Statistic	Prob.
ADF	-4.659987	0.0000
Residual variance	0.001989	
HAC variance	0.003202	

4. Functional form and capacity constraint

Functional form should account for the fact that the traffic rate of growth diminishes as it approaches the capacity limit.

Partial adjustment model, with variable adjustment speed:

$$\Delta \ln Y_{it} = \lambda_{it} \cdot (\ln Y*_{it} - \ln Y_{it-1}) + \varepsilon_{it}$$

$$\ln Y*_{it} = \alpha_i + \beta \cdot \ln X_{it}$$

4. Functional form and capacity constraint

We define the level of quality of the motorway in terms of traffic volume:

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

Our assumption is that the speed of adjustment is a function of the level of quality:

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

4. Functional form and capacity constraint

We can consider two polar cases

$$Y_{it-1} = Y_i^0 \Rightarrow \tau_{it} = 0 \Rightarrow \lambda_{it} = 0$$

$$Y_{it-1} = 0 \Rightarrow \tau_{it} = 1 \Rightarrow \lambda_{it} = \theta$$

By substituting the expression for y* in the dynamic equation and λ we obtain the following equations

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

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

4. Functional form and capacity constraint

The previous formulation can be generalised to account for s lags

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

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

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

4. Functional form and capacity constraint

By substituing Y* and λ we have:

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

That is,

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

5. Estimated equation

Dependent Variable: D(Ln(traffic))/tau

Estimation method: weighted least squares

	Coefficient	Std. Error	t-Statistic
ln(GDP)	0.801667	0.04342	18.461
ln(Pgas)	-0.384932	0.01698	-22.674
ln(traffic(-1))	-0.635306	0.02413	-26.329
ln(toll1)	-0.158214	0.01740	-9.094
ln(toll2)	-0.336049	0.02146	-15.659
ln(toll3)	-0.495918	0.03041	-16.305
AR(1)	0.751514	0.02157	34.841
Dummy variables			
Fixed effects			
R^2	0.60		
Observations	1694		

6. Elasticities

Elasticities depend on the level of traffic flow

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

Can be rewritten as:

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

$$\ln Y_{it} = \beta^*_{0it} + \beta^*_{1it} \ln X_{it} + \gamma^*_{it} \ln Y_{t-1}$$

Where,

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

6. Elasticities

For a given level of traffic flow τ we have

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

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

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

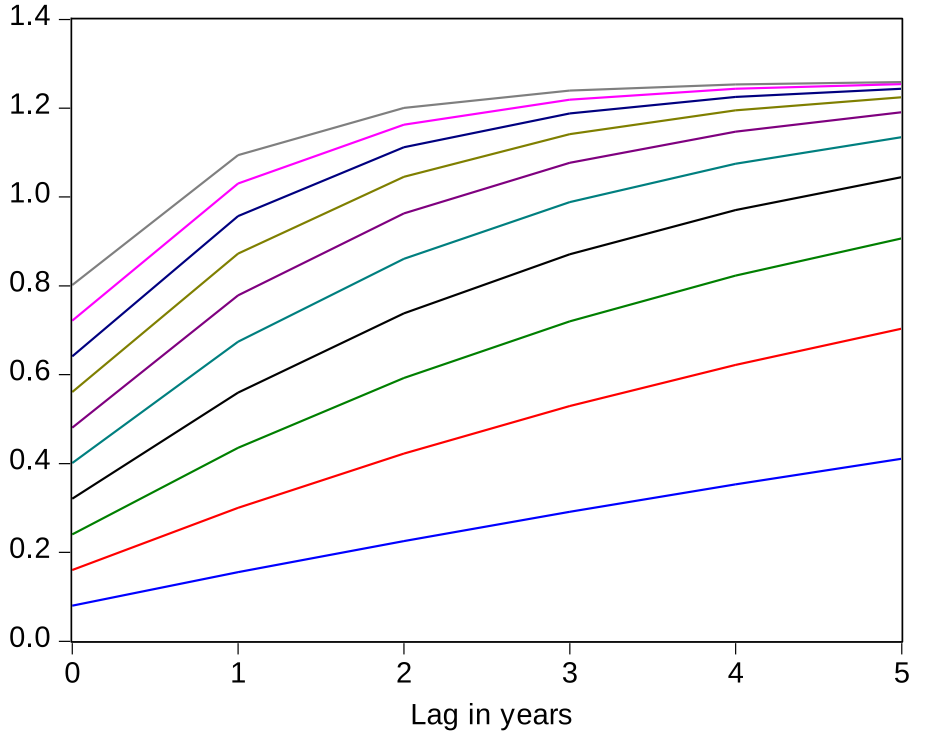
$$\varepsilon_J = \beta_1^* \cdot \frac{(1 - \gamma^{*J+1})}{(1 - \gamma^*)}$$

6. Elasticities

Elasticity with respect to GDP (different values for τ)

J (years)	tau=0.1	tau=0.5	tau=1
0	0.0802	0.4009	0.8017
1	0.1552	0.6744	1.0941
2	0.2256	0.8610	1.2007
3	0.2914	0.9884	1.2396
4	0.3531	1.0753	1.2538
5	0.4108	1.1346	1.2590

6. Elasticities



6. Elasticities

For the specific case $\tau_{it} = 1$

Short term: $\varepsilon_{J=0} = \beta_1^* = \beta_1$

Long term: $\varepsilon_{\infty} = \dfrac{\beta_1^*}{1-\gamma^*} = \dfrac{\beta_1}{\theta}$

6. Elasticities

Elasticities for $\tau_{it} = 1$

	Short term	Long term
GDP	0.802	1.262
Gasoline price	-0.385	-0.606
Toll 1	-0.158	-0.249
Toll 2	-0.336	-0.529
Toll 3	-0.496	-0.781

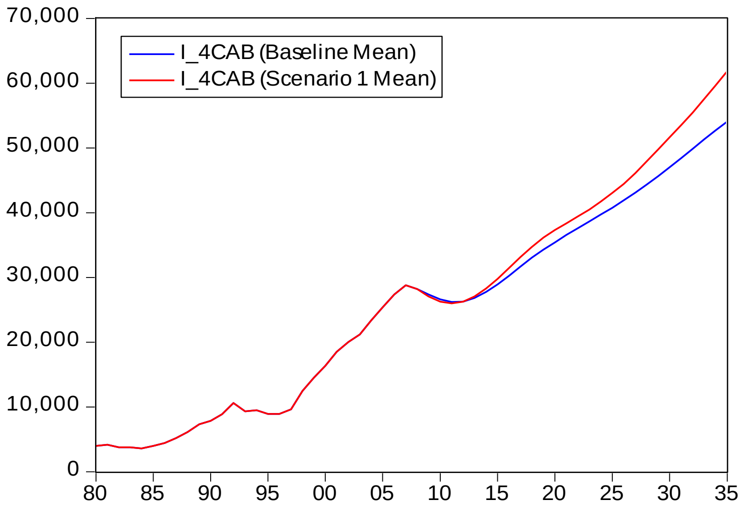
7. Uncertainty in traffic forecast

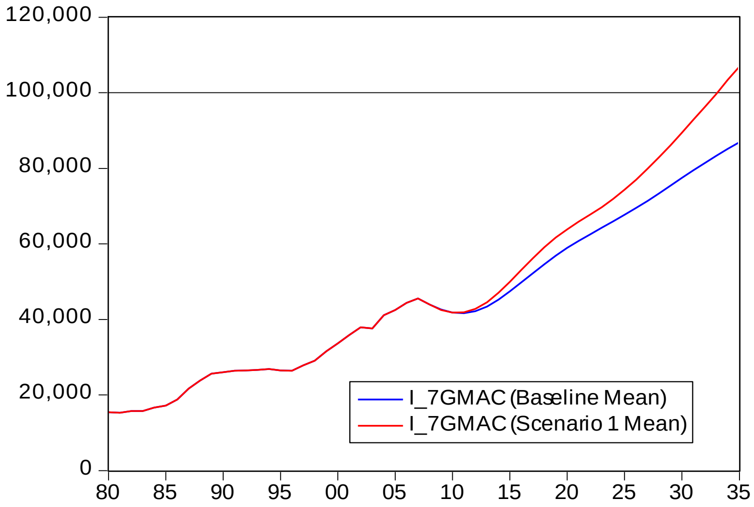
Forecasting period: 2009-2035

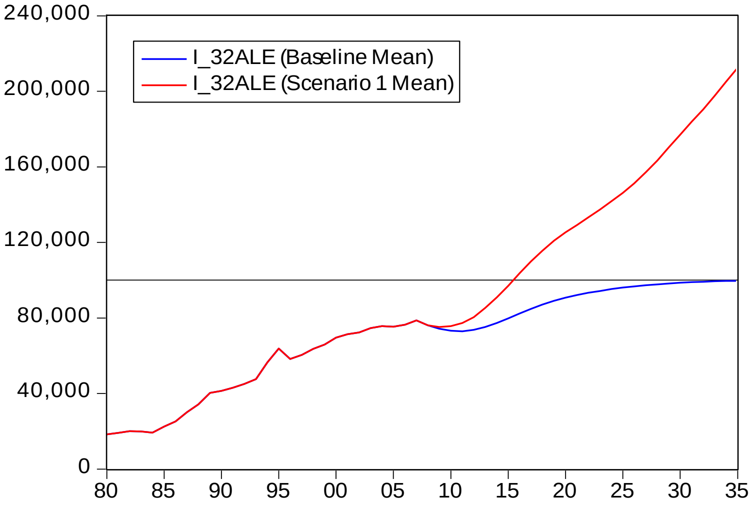
Explanatory variables:

- Motorway toll is assumed to remain constant
- GDP: ARIMA mode
- Gasoline Price: ARIMA model

7. Traffic forecast for three representative sections







7. Uncertainty in traffic forecast

Uncertainty in the forecast can be due to:

- Input uncertainty: the future values of exogenous variables are unknown
- Model uncertainty:
  - Random term uncertainty
  - Coefficient uncertainty

7. Uncertainty in traffic forecast

Random term uncertainty is obtained through a stochastic simulation process. The model is solved repeatedly for different draws of random terms of the model.

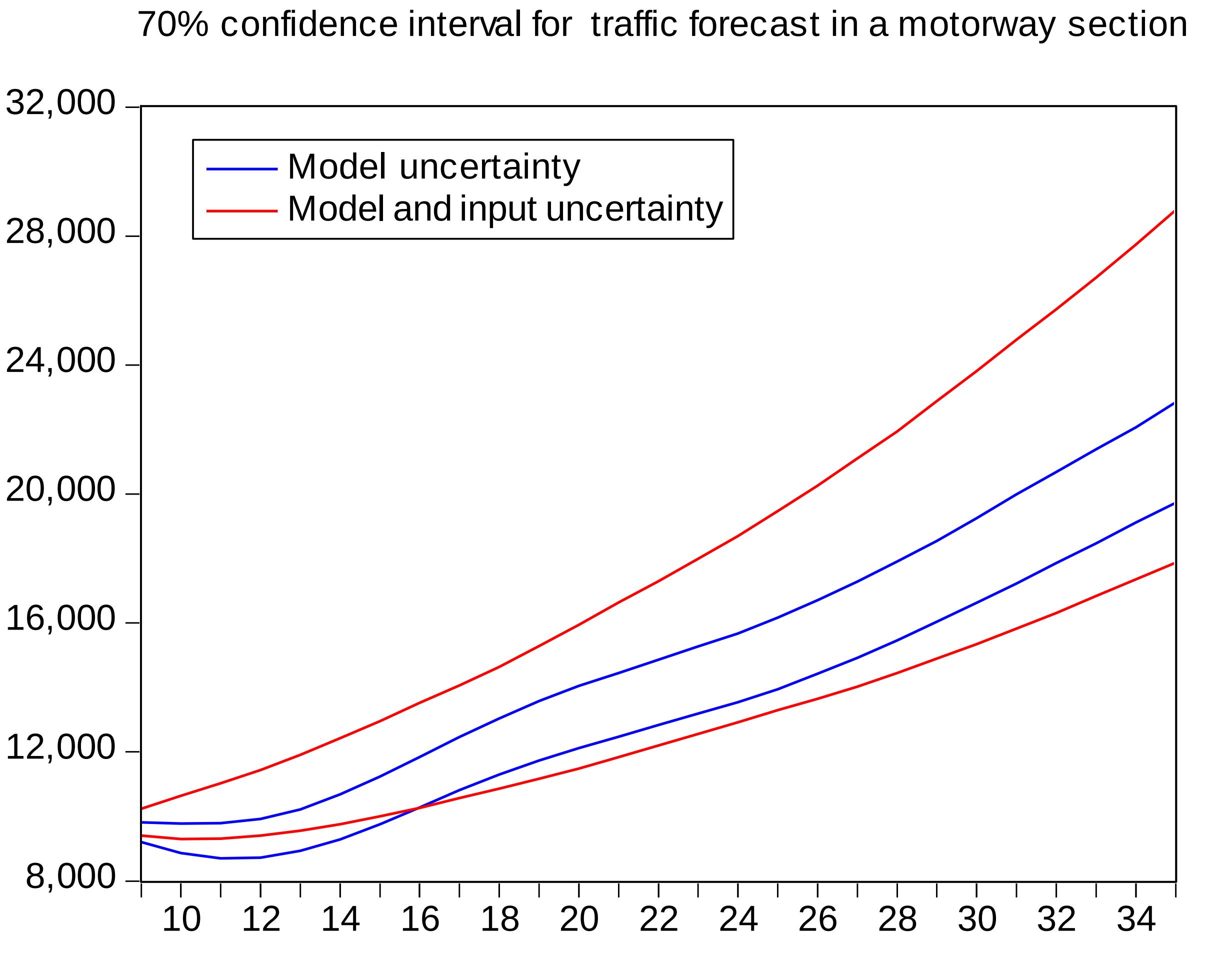
Coefficient uncertainty is included in the model; a new set of coefficients is drawn before each repetition.

Input uncertainty is included through a simulation process; a new set of values is drawn before each repetition.

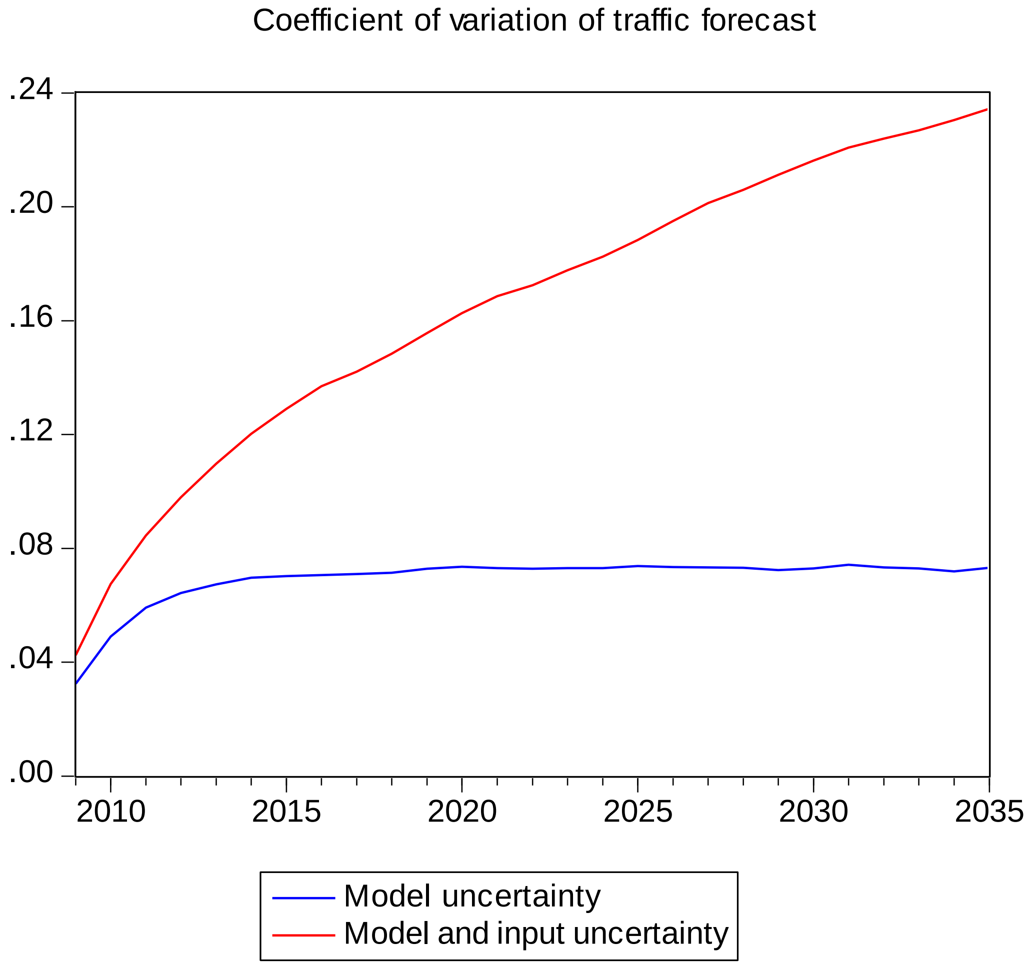
7. Uncertainty in traffic forecast

Once the model has been solved many times using different draws for the random components, the statistics over all the different outcomes are calculated: mean and standard deviation.

7. Uncertainty in traffic forecast



7. Uncertainty in traffic forecast



7. Uncertainty in traffic forecast

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

		Model	Inputs
2009	0.0424	76.4%	23.6%
2010	0.0675	72.7%	27.3%
2011	0.0845	70.0%	30.0%
2012	0.0979	65.7%	34.3%
2013	0.1097	61.3%	38.7%
2014	0.1202	58.0%	42.0%
2015	0.1290	54.5%	45.5%
2016	0.1369	51.6%	48.4%
2017	0.1420	50.0%	50.0%
2018	0.1484	48.1%	51.9%
2019	0.1556	46.8%	53.2%
2020	0.1626	45.2%	54.8%
2021	0.1686	43.3%	56.7%
2022	0.1725	42.2%	57.8%
2023	0.1777	41.1%	58.9%
2024	0.1825	40.0%	60.0%
2025	0.1883	39.2%	60.8%
2026	0.1950	37.6%	62.4%
2027	0.2012	36.4%	63.6%
2028	0.2059	35.5%	64.5%
2029	0.2113	34.3%	65.7%
2030	0.2163	33.7%	66.3%
2031	0.2208	33.6%	66.4%
2032	0.2240	32.7%	67.3%
2033	0.2269	32.1%	67.9%
2034	0.2305	31.2%	68.8%
2035	0.2344	31.2%	68.8%

7.Uncertainty and the expected value of traffic

In a regression model

$$Y_{T+J} = \exp(\beta \cdot X_{T+J} + u_{T+J})$$

$$X_{T+J} = X_T(J) + v_{T+J}$$

The value of the dependent variable in period T+J is:

$$Y_{T+J} = \exp\left[\beta \cdot X_T(J) + u_{T+J} + v_{T+J}\right] = \exp\left[\beta \cdot X_T(J)\right] \cdot \exp(u_{T+J}) \cdot \exp(v_{T+J})$$

Under normality hypothesis the expected value is:

$$E\left[Y_{T+J}\right] = \exp\left[\beta \cdot X_T(J)\right] \cdot \exp\left[\frac{\sigma_u^2}{2}\right] \cdot \exp\left[\frac{\sigma_v^2}{2}\right]$$

7.Uncertainty and the expected value of traffic

Deterministic simulation

$$\exp\left[\beta \cdot X_T(J)\right]$$

Simulation with model uncertainty (random term and coefficients)

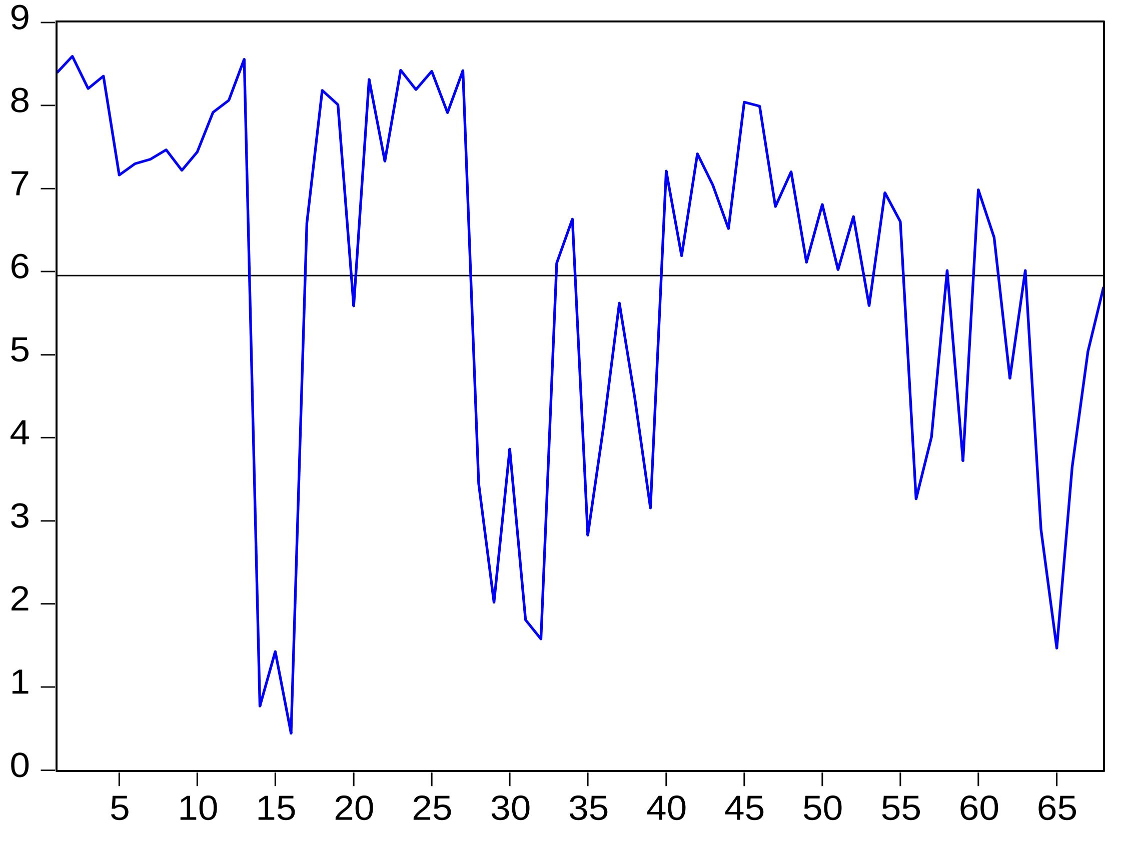
$$\exp\left[\beta \cdot X_T(J)\right] \cdot \exp\left[\frac{\sigma_u^2}{2}\right]$$

Simulation with model and input uncertainty

$$\exp\left[\beta \cdot X_T(J)\right] \cdot \exp\left[\frac{\sigma_u^2}{2}\right] \cdot \exp\left[\frac{\sigma_v^2}{2}\right]$$

7.Uncertainty and the expected value of traffic

Percentage of difference in the average value of traffic forecast: deterministic vs stochastic (2009-2035)



8. Conclusions

1. The introduction of a capacity constraint in the demand equation significantly affects the traffic forecast as the volume of traffic increases.
2. To obtain confidence intervals, input uncertainty should be taken into account in long term forecasting
3. In non linear models, stochastic simulations are necessary to obtain unbiased forecast