Measuring uncertainty of traffic volume on motorway concessions: a time-series analysis

Antonio Sánchez Solño*, Antonio L. Lara Galera, Fernando Cabero Colín

*Universidad Politécnica de Madrid, c/ Profesor Aranguren s/n, Madrid 28040, Spain

Abstract

This paper presents a method to build up a confidence interval for the evolution of traffic in motorway concessions, based on a univariate time-series model. The main advantage of this method, compared to traditional traffic models, is that it allows to avoid the error in the prediction of the explanatory variables. The results obtained show that the use of a time-series model represents a feasible alternative to assess traffic uncertainty in existing concessions, when long series of traffic data are available.

© 2017 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 20th EURO Working Group on Transportation Meeting.

Keywords: traffic; motorway; time-series; unit root; concession

1. Introduction

Traditionally, the first steps in long-term traffic models are based on the establishment of a relationship between transportation demand and certain explanatory variables for which available information and prediction capacity are greater than for traffic itself. Then, in successive steps, transportation mode choice and route choice models are applied (Department of Transport, 2006).

However, in certain cases (as discussed below) one could expect that the evolution of traffic itself contributes better information than other variables. Then, the use of univariate time-series models may be an alternative tool to predict the traffic volume and to build a confidence interval for the forecast, when there are available data for traffic during a long enough period.

* Corresponding author. Tel.: +34-91-3366649
E-mail address: asanoli@ciicp.es
In any case, building a confidence interval for traffic forecasts is a complex issue in traditional transportation models. As De Jong et al. (2007) and Matas et al. (2012) state, the literature on quantifying uncertainty in traffic forecasts is fairly limited. In most traffic studies, the issue of uncertainty is addressed by considering different scenarios for the exogenous variables (for example, different rates of economic growth). However, this procedure does not really quantify uncertainty if it does not provide the likelihood of each alternative forecast (Matas et al., 2012). Therefore, statistical methods are needed to quantify uncertainty.

A statistical analysis of traffic uncertainty should consider both the uncertainty due to model inputs and the uncertainty due to the model itself. The uncertainty associated with model inputs can be measured by estimating the probability distributions of exogenous variables. From these distributions can be obtained, using an analytical method or a simulation process, the contribution of the uncertainty in the inputs to the uncertainty in predicting traffic (Boyce and Bright, 2003). To this must be added the model specification errors and errors in the determination of the parameters of the model (Brundell-Freij, 2000; Beser Hugosson, 2005).

The work of De Jong et al. (2007) obtained as one of its main conclusions that the uncertainty due to the inputs of the model is much more important than the uncertainty due to the model itself. This may lead to the reflection that, when there is appropriate information on the evolution of traffic in a relatively stable environment, it may be more useful an analysis based on traffic time-series that the development of a complex traffic model.

Possibly, the choice of a univariate time-series model is not suitable for the evaluation of a new road infrastructure project, but it may be appropriate in other cases, for example, the appraisal of an existing motorway concession. In this regard, it is noteworthy that one could expect future developments of secondary markets of infrastructure concessions (Alcaraz and Sánchez Soliño, 2015), which in turn will require the development of analytical tools for making proper assessments of traffic in such concessions.

2. Unit root analysis

A first step in the characterization of time-series is the analysis of stationarity. In order to perform a unit root (or non-stationarity) test of time-series, we start from an autoregressive model that can be expressed as:

\[ y_t = \alpha + \rho y_{t-1} + \varepsilon_t \]  

where:
- \( y_t \): random variable
- \( \alpha \): intercept (constant)
- \( \rho \): constant parameter
- \( \varepsilon_t \): white noise

Subtracting the term \( y_{t-1} \) from both sides of equation (1), we obtain:

\[ y_t - y_{t-1} = \alpha + (\rho - 1)y_{t-1} + \varepsilon_t \]  

In order to make the calculation easier, we can say \( \beta = \rho - 1 \), so equation (2) can be written as follows:

\[ y_t - y_{t-1} = \alpha + \beta y_{t-1} + \varepsilon_t \]  

Then, we could try to estimate the parameter \( \beta \) by using Ordinary Least Squares (OLS), and calculating the t-statistic to test whether \( \beta \) is significantly different from 0. If we cannot reject the hypothesis that \( \beta = 0 \), then we can say that the process has a unit root, and cannot reject that the \( y_t \) variable is non-stationary. However, if the true value of \( \rho \) is 1 (\( \beta = 0 \)), then the OLS estimator of \( \rho \) is biased towards zero (Pindyck & Rubinfeld, 1998). Then the use of OLS could lead us to incorrectly reject the non-stationarity hypothesis.

To solve this problem, Dickey-Fuller (1979, 1981) used a Monte Carlo simulation to calculate the correct critical values for the distribution of the t-statistic when \( \rho = 1 \). Additionally, other authors have obtained these critical values, such as McKinnon (McKinnon, 1990, 2010). If the t-statistic obtained in our estimation is greater than the
critical value, we cannot reject the null hypothesis $\beta = 0$, and we cannot reject that the traffic time-series is non-stationary.

In this kind of test we assume that there is no serial correlation in the error term $\varepsilon_t$. However, the process described before could be non-stationary, even when serial correlation exists. So, the same authors (Dickey-Fuller 1979, 1981), proposed an extended method which is known as Augmented Dickey-Fuller test (ADF). In this test, the model is expanded by adding the lagged dependent variable to the right side of the equation, as follows:

$$\Delta y_t = \alpha + \beta y_{t-1} + \sum_{j=1}^{m} \lambda_j \Delta y_{t-j} + \varepsilon_t$$

(4)

where $\lambda_j$ represents the $m$ parameters obtained in the regression analysis between the dependent variable $\Delta y_t$ and the same dependent variable with a lag of $j$ periods. The number of lags taken in this paper is the minimum compatible with the absence of serial correlation in the residuals.

The regression analysis to determine the parameters in equation 4 is made using OLS. The $t$-statistic obtained for the parameter $\beta$ is then compared with the same critical values contained in the former Table 1. Again, if the $t$-statistic obtained in our estimation is greater than the critical value, we cannot reject that $\beta = 0$ and cannot reject that the process is non-stationary.

We have then applied this procedure to the main toll motorways in Spain. Data were taken from the public information supplied by the concessionaire firms, and published by the Spanish Ministry of Public Works (2015). For this research we used the Annual Average Daily Traffic (AADT) in each motorway in order to avoid the seasonality problem in traffic volumes.

In order to perform the ADF test, we considered: $y_t = \ln(\theta_t)$ and $\Delta y_t = \ln(\theta_t/\theta_{t-1})$, where $\theta_t$ is the traffic volume in the year $t$ in terms of AADT and $\Delta y_t$ represents the continuous growth rate of traffic. Then, equation (4) takes the form:

$$\ln \left( \frac{\theta_t}{\theta_{t-1}} \right) = \alpha + \beta \ln(\theta_{t-1}) + \sum_{j=1}^{m} \lambda_j \ln \left( \frac{\theta_{t-j}}{\theta_{t-j-1}} \right) + \varepsilon_t$$

(5)

Starting from this equation, we applied a regression analysis, using OLS to obtain the estimation of the parameter $\beta$ and the $t$-statistic for that estimation.

Out of the twenty seven toll motorways existing in Spain, we analyzed only those where traffic time-series are long enough (more than thirty years). The obtained results are given in Table 1.

<table>
<thead>
<tr>
<th>Motorway</th>
<th>Period</th>
<th>Number of lags</th>
<th>ADF test t-statistic</th>
<th>5% critical value</th>
<th>Durbin-Watson statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sevilla – Cádiz</td>
<td>1974 – 2014</td>
<td>1</td>
<td>-0.9034</td>
<td>-2.9369</td>
<td>1.8213</td>
</tr>
<tr>
<td>Bilbao – Zaragoza</td>
<td>1978 – 2014</td>
<td>1</td>
<td>-1.4172</td>
<td>-2.9484</td>
<td>1.8658</td>
</tr>
<tr>
<td>Burgos – Armiñón</td>
<td>1978 – 2014</td>
<td>1</td>
<td>-1.8945</td>
<td>-2.9484</td>
<td>2.0093</td>
</tr>
<tr>
<td>León – Campomanes</td>
<td>1983 – 2014</td>
<td>1</td>
<td>-0.8449</td>
<td>-2.9640</td>
<td>1.7742</td>
</tr>
<tr>
<td>Tarragona – Valencia</td>
<td>1974 - 2014</td>
<td>1</td>
<td>-1.8029</td>
<td>-2.9369</td>
<td>1.7365</td>
</tr>
<tr>
<td>Ferrol – Portugal</td>
<td>1982 – 2014</td>
<td>1</td>
<td>-1.7816</td>
<td>-2.9640</td>
<td>1.8808</td>
</tr>
</tbody>
</table>
As can be observed, the result of the analysis (taking into account the critical values for a 5% significance level) is that we cannot reject the null hypothesis of the existence of a unit root in any of the toll motorways. Moreover, the Durbin-Watson test shows that the serial correlation has been eliminated in all cases after considering only one lag.

3. Calibration of a traffic time-series model

Taking into account the results of the latter section, we can assume that β = 0 in equation (5). In the case of one lag, we would have:

\[ \Delta y_t = \alpha + \lambda \Delta y_{t-1} + \varepsilon_t \]  

(6)

Where \( \Delta y_t = \ln\left(\frac{\theta_t}{\theta_{t-1}}\right) \) and \( \Delta y_{t-1} = \ln\left(\frac{\theta_{t-1}}{\theta_{t-2}}\right) \)

This would mean that the growth rate of traffic follows an autoregressive model AR (1), while traffic itself would follow an homogeneous non-stationary process of order 1. In this case, it is possible to estimate the parameters \( \alpha \) and \( \lambda \) in equation (6) by applying a linear regression, starting from the real data in a specific motorway (Pindyck and Rubinfeld, 1998).

As an example, we have taken the case of the Villalba (Madrid)-Adanero motorway, where real traffic can be seen in table 2.

Table 2. Traffic volume in the Villalba-Adanero motorway (Annual Average Daily Traffic)

<table>
<thead>
<tr>
<th>Year</th>
<th>Traffic volume (θt)</th>
<th>Growth rate (ln θt/θt-1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974</td>
<td>7.258</td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td>7.817</td>
<td>0.07420</td>
</tr>
<tr>
<td>1976</td>
<td>8.168</td>
<td>0.04392</td>
</tr>
<tr>
<td>1977</td>
<td>6.690</td>
<td>-0.19961</td>
</tr>
<tr>
<td>1978</td>
<td>7.796</td>
<td>0.15300</td>
</tr>
<tr>
<td>1979</td>
<td>8.455</td>
<td>0.08115</td>
</tr>
<tr>
<td>1980</td>
<td>8.326</td>
<td>-0.01537</td>
</tr>
<tr>
<td>1981</td>
<td>8.380</td>
<td>0.00646</td>
</tr>
<tr>
<td>1982</td>
<td>8.355</td>
<td>-0.00299</td>
</tr>
<tr>
<td>1983</td>
<td>8.283</td>
<td>-0.00865</td>
</tr>
<tr>
<td>1984</td>
<td>8.452</td>
<td>0.02020</td>
</tr>
<tr>
<td>1985</td>
<td>8.810</td>
<td>0.04148</td>
</tr>
<tr>
<td>1986</td>
<td>9.478</td>
<td>0.07309</td>
</tr>
<tr>
<td>1987</td>
<td>10.360</td>
<td>0.08898</td>
</tr>
<tr>
<td>1988</td>
<td>11.420</td>
<td>0.09741</td>
</tr>
<tr>
<td>1989</td>
<td>12.929</td>
<td>0.12411</td>
</tr>
<tr>
<td>1990</td>
<td>14.005</td>
<td>0.07994</td>
</tr>
<tr>
<td>1991</td>
<td>15.610</td>
<td>0.10850</td>
</tr>
<tr>
<td>1992</td>
<td>16.415</td>
<td>0.05028</td>
</tr>
<tr>
<td>1993</td>
<td>16.504</td>
<td>0.00541</td>
</tr>
<tr>
<td>1994</td>
<td>16.628</td>
<td>0.00749</td>
</tr>
</tbody>
</table>
For this series, and using Ordinary Least Squares (OLS), we obtain the value of the parameters of the autoregressive model:

\[ \alpha = 0.0221 \]
\[ \lambda = 0.2844 \]

The value of the variance of the residuals \( \varepsilon_t \) is, in this case:

\[ \sigma^2 = 0.0041 \]

Observe that the value of the mean of the series \( \Delta y_t \) is given by:

\[ \mu = \frac{\alpha}{1-\lambda} \]  

Starting from these results, we can perform a forecast for the time-series of the growth rate of traffic in the motorway. In a general case, if the realized time-series includes \( T \) periods (years), the one-period forecast will be (we will call \( w_t = \Delta y_t \)):

\[ w_{T+1}^* = E [w_{T+1} \mid w_T, w_{T-1}, ..., w_1] = \alpha + \lambda w_T \]  

Where \( E \) represents the expected value operator. Analogously, the two-period forecast will be:

\[ w_{T+2}^* = \alpha + \lambda w_{T+1}^* = \alpha (1 + \lambda) + \lambda^2 w_T \]  

And the \( n \)-period forecast will be:

\[ w_{T+n}^* = \alpha (\lambda^{n-1} + \lambda^{n-2} + ... + \lambda + 1) + \lambda^n w_T \]
For large values of n, it is easy to demonstrate that the forecast converges to the mean of the series \( \mu \). In figure 1, we can observe the forecast for years 2015-2019 in our example. In this case, the convergence to the mean takes place in a very few periods.

In this model, the forecast error \( e \) will be:

For one period:

\[
e_w(1) = w_{T+1} - w_{T+1}^* = \alpha + \lambda w_T + \varepsilon_{T+1} - (\alpha + \lambda w_T) = \varepsilon_{T+1}
\]  

(11)

For two periods:

\[
w_{T+2} = \alpha + \lambda w_{T+1} + \varepsilon_{T+2} = \alpha + \lambda (\alpha + \lambda w_T + \varepsilon_{T+1}) + \varepsilon_{T+2} = \alpha (1+\lambda) + \lambda^2 w_T + \lambda \varepsilon_{T+1} + \varepsilon_{T+2}
\]

\[
w_{T+2}^* = \alpha (1+\lambda) + \lambda^2 w_T
\]

\[
e_w(2) = w_{T+2} - w_{T+2}^* = \varepsilon_{T+2} + \lambda \varepsilon_{T+1}
\]  

(12)

And, similarly, for n periods:

\[
e_w(n) = w_{T+n} - w_{T+n}^* = \varepsilon_{T+n} + \lambda \varepsilon_{T+n-1} + \ldots + \lambda^{n-1} \varepsilon_{T+1}
\]  

(13)

The variance of this forecast error will be:

\[
V[e_w(n)] = E[e^2(n)] = (1 + \lambda^2 + \lambda^4 + \ldots + \lambda^{2n-2}) \sigma_e^2
\]  

(14)

For traffic itself (in fact, for the logarithmic transformation of traffic \( y_t = \ln \theta_t \)), forecasts will be related to forecasts of the differenced series (Pindyck and Rubinfeld, 1998).

Then, the one-period forecast error for the logarithm of traffic will be:

\[
e(1) = y_{T+1} - y_{T+1}^* = y_T + w_{T+1} - y_T - w_{T+1}^* = \varepsilon_{T+1}
\]  

(15)
The two-period forecast error will be:

\[ e(2) = y_{T+2} - y_{T+2}^* = y_T + w_{T+1} + w_{T+2} - y_T - w_{T+1}^* - w_{T+2}^* = (1 + \lambda) \epsilon_{T+1} + \epsilon_{T+2} \]  

(16)

And similarly, for \( n \) periods:

\[ e(n) = (1 + \lambda + \lambda^2 + \ldots + \lambda^{n-1}) \epsilon_{T+1} + (1 + \lambda + \ldots + \lambda^{n-2}) \epsilon_{T+2} + \ldots + (1 + \lambda) \epsilon_{T+n-1} + \epsilon_{T+n} \]  

(17)

The variances of these forecast errors are the following:

\[ E[e^2(1)] = \sigma^2 \]  

(18)

\[ E[e^2(2)] = \sigma^2 \left[ 1 + (1 + \lambda)^2 \right] \]  

(19)

\[ \ldots \]  

\[ E[e^2(n)] = \sigma^2 \sum_{i=1}^{n} (\lambda^0 + \lambda^1 + \lambda^2 + \ldots + \lambda^{n-1})^2 \]  

(20)

Observe that a random walk model could be treated as a particular case, in which the value of the parameter \( \lambda \) is equal to zero. In that case, it is immediately deducted that the variance of the error forecast for \( n \) periods is equal to \( n\sigma^2 \).

In our example, we would obtain:

\[ E[e^2(1)] = 0.0041 \]

\[ E[e^2(2)] = 0.0041[1 + (1 + 0.2844)^2 + 1] = 0.0109 \]

\[ E[e^2(3)] = 0.0041[(1 + 0.2844 + 0.2844^2)^2 + (1 + 0.2844)^2 + 1] = 0.0185 \]

\[ E[e^2(4)] = 0.0264 \]

\[ E[e^2(5)] = 0.0344 \]

\[ \ldots \]

As can be seen, the variance of the forecast error grows when \( n \) is greater. This calculation allows us to build up a confidence interval for the forecast given by our model. For example, we would obtain a 95% confidence interval by considering \( \pm 1.96 \) times the standard deviation of the forecast error. In figure 2, we have plotted this confidence interval for the forecast of the traffic volume.

Fig. 2. 95% confidence interval for (the natural logarithm of) traffic volume (AADT) in the Villalba-Adanero motorway
In this way, we can obtain a measure of the uncertainty involved in the prediction of traffic in this concession project. As can be seen, an important result is that the confidence interval obtained grows with time and with the parameter \( \lambda \) that defines the autoregressive character of the traffic growth rate.

4. Conclusions

In motorway concessions, both the forecast of the traffic volume and the measure of the risk involved are essential for the appraisal of the business. So far, the focus on this issue has been put on newly built motorway projects; however, an increase in transactions over existing concessions, where there is a history of traffic over a number of years, is increasingly expected. At the same time, the measure of traffic uncertainty constitutes a fundamental input for the evaluation of the project from the point of view of the Public Administration, since the latter can consider a new sale of the concession once the motorway returns to the public sector, at the end of the concession period.

In these cases, the outcome obtained in this paper is that the use of a univariate time-series model represents a feasible alternative to build up a confidence interval for the prediction of traffic volume, as a tool to measure the risk of the project. The analysis carried out is based on data from traffic series of more than thirty years on Spanish toll motorways. The results show that the non-stationarity of the traffic series cannot be rejected for any of the analyzed projects. Additionally, a procedure is shown to obtain a confidence interval, based on the autoregressive nature of the traffic growth rate.

This type of model has the advantage of avoiding the error in the prediction of the explanatory variables themselves. In this context, the use of this methodology, instead of traditional transportation models, makes sense when there are real traffic data over an extended period.

Acknowledgements

This work was carried out within the research project TRA2015-64723-R (MINECO/FEDER), financed by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund.

References


