UNIT ROOT ANALYSIS OF TRAFFIC TIME SERIES IN TOLL HIGHWAYS

Antonio Sánchez Soliño, Associate Professor, Universidad Politécnica de Madrid
E-mail: asanoli@ciccp.es

Antonio L. Lara Galera, Associate Professor, Universidad Politécnica de Madrid
E-mail: alargal@ciccp.es

ABSTRACT

Often, concession contracts in highways include some kind of clauses (for example, a minimum traffic guarantee) that allow for a better management of the business risks. The value of these clauses may be important and should be added to the total value of the concession. However, in these cases, traditional valuation techniques, like the Net Present Value (NPV) of the project, are insufficient. An alternative methodology for the valuation of highway concession is the one based on the real options approach. This methodology is generally built on the assumption of the evolution of traffic volume as a Geometric Brownian Motion (GBM), which is the hypothesis analyzed in this paper.

In this paper, we describe first the methodology used for the analysis of the existence of unit roots (i.e., the hypothesis of non-stationarity) in time series in general. We have used the Dickey-Fuller approach, which is the most widely used test for this kind of analysis. Then we apply this methodology to perform a statistical analysis of traffic series in Spanish toll highways. For that purpose, we use data on the Annual Average Daily Traffic (AADT) in a set of highways. The period of analysis is around thirty years in most cases.

The main outcome of the research is that we cannot reject the hypothesis that traffic volume follows a GBM process in Spanish toll highways. This result is robust, and therefore we can use it as a starting point for the application of the real options theory to valuate toll highway concessions.

INTRODUCTION

Most of road traffic models are based on the relationship between traffic volume and a number of explicative variables for which available information and prediction capacity are greater than for traffic itself. However, the use of 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 in a given road during a long enough period.

In this case, we can assume, in principle, that the evolution of traffic volume follows a Geometric Brownian Motion (GBM), which can be described in the following way:

$$d\theta = a\theta\, dt + \sigma\theta\, dz \tag{1}$$

where:

θ : traffic volume
$d\theta$: differential increment of traffic
a : growth rate of traffic.
dt : differential time interval
σ : traffic volatility
dz : increment of a Wiener process

Starting from equation (1), and applying Itô´s lemma, we can find the process followed by the natural logarithm of θ (Itô, 1951):

$$d(\ln\theta) = a\acute{}\, dt + \sigma\, dz \tag{2}$$

where ***ln θ*** is the natural logarithm of traffic and $a\acute{} = a - \sigma^2/2$.

On the right-hand side of equation 2, the parameter ***a´*** is a constant drift term or growth parameter. It means that the logarithm of traffic has a growth of *a´* per unit of time. Regarding the second term, ***dz*** is the increment of a standard Wiener process, so that $dz=\varepsilon_t(dt)^{1/2}$, where $\boldsymbol{\varepsilon_t}$ is a variable which is normally distributed with zero mean and unit standard deviation (Dixit and Pindyck, 1994). This second term, ***σdz***, adds a noise or variability to the path followed by the logarithm of traffic. The amount of this noise is ***σ*** times a standard Wiener process, so the process represented by equation 2 has a standard deviation of ***σ***. This means that the variance rate (the variance per unit of time) of this process is $\boldsymbol{\sigma^2}$ (Hull, 2006). We assume that the parameter ***σ,*** which is called the traffic volatility, is also a constant.

The discrete version of equation (2) would be the following:

$$\Delta(\ln\theta) = a\acute{}\,\Delta t + \sigma\,\Delta z \tag{3}$$

where:

- E(Δz) = 0 [expected value of Δz]
- E[$\Delta(\ln\theta)$] = $a\acute{}\,\Delta t$ [expected value of $\Delta(\ln\theta)$]
- V[$\Delta(\ln\theta)$] = $\sigma^2\Delta t$ [variance of $\Delta(\ln\theta)$]

This means that the change in the logarithm of traffic is normally distributed over any time interval Δt (with mean $a'\Delta t$ and standard deviation $\sigma \sqrt{\Delta t}$), following a random walk with a drift. This assumption is frequently made for economic and financial variables. For stock prices, for example, the hypothesis of GBM is generally accepted, and it has been used for the development of the theory of options´ valuation, since the initial works carried out by Black and Scholes (1973) and Merton (1973). In the field of road traffic, this assumption has been made by Zhao *et al.* (2004) to analyze the decision-making process in highway development.

However, the GBM hypothesis is not always evident. Pyndick and Rubinfeld (1998), for example, have analyzed if commodity prices follow this process. They found that, for very long time series (more than 100 years), detrended prices of crude oil and copper do not follow a random walk, but a mean-reverting process instead. In the contrary case, the hypothesis of a random walk cannot be rejected for the detrended prices of lumber.

In this paper, we perform a test for the hypothesis of a GBM for the evolution of traffic volume in toll highways. We have used the series available for Spanish toll highways, which, in most cases, cover a thirty years period. First, we describe in the next section the methodology used for the analysis of the existence of unit roots in time series in general. We have used the Dickey-Fuller approach, which is the most widely used test for this kind of analysis. Then we apply this methodology for traffic series in Spanish toll highways and we discuss the results obtained. We consider the limitations of the analysis carried out and we discuss the possible application of the results. Finally, we resume the main conclusions of the paper.

UNIT ROOTS ANALYSIS OF TIME SERIES

Let us suppose that we have a random variable Y_t which evolves over time following an autoregressive process that can be described as:

$$Y_t = \rho Y_{t-1} + u_t \qquad (4)$$

where u_t is a random error term. Now, we would like to analyze the parameter ρ. If ρ is equal to 1, then it is said that a unit root exists, which means that Y_t is a non-stationary variable. In the contrary case (if $\rho \neq 1$) then the variable Y_t would be stationary.

We can add a constant drift term α to equation 4, without changing the reasoning. Then, the equation would be:

$$Y_t = \alpha + \rho Y_{t-1} + u_t \qquad (5)$$

We can rewrite equation 5 in the following way:

$$Y_t - Y_{t-1} = \alpha + (\rho - 1) Y_{t-1} + u_t \qquad (6)$$

We could try to estimate the parameter ρ in equation 6 using Ordinary Least Squares (OLS), and calculating the t-statistic to test whether ρ is significantly different from 1. If we cannot reject the hypothesis that $\rho = 1$, then we say that the process has a unit root, and we cannot reject that the variable Y_t is non-stationary after detrending. However, if the true value of ρ is

1, then the OLS estimator is biased toward zero (Pindyck and Rubinfeld, 1998). Then the use of OLS could lead us to incorrectly reject the non-stationarity hypothesis.

To solve this problem, Dickey and Fuller (1979, 1981) used a Monte Carlo simulation to calculate the correct critical values for the distribution of the t-statistic when $\rho = 1$. Thus the Dickey-Fuller (DF) test is the most widely used one to analyze the existence of a unit root in a given process.

To apply the DF test, we write equation 6 in the following way:

$$\Delta Y_t = \alpha + \beta Y_{t-1} + u_t \qquad (7)$$

where $\beta = \rho - 1$

Now, we perform the OLS method to estimate the value of the parameter β (where the null hypothesis is that $\beta = 0$) and to calculate its t-statistic. Then, we compare the t-statistic thus obtained with the critical values calculated by Dickey-Fuller. In fact, we will use the critical values obtained by other authors based on the DF methodology. For example, McKinnon (1990) obtained the following critical values.

Table 1 - Critical values for t-statistic in DF unit roots tests

Sample size	Significance level = 5%	Significance level = 10%
25	-3.00	-2.63
50	-2.93	-2.60
100	-2.89	-2.58
∞	-2.86	-2.57

If the t-statistic obtained in our estimation is greater than the critical value, then we cannot reject that $\beta = 0$ and, therefore, we cannot reject that the process is non-stationary after detrending. Observe that all critical values are negative. Therefore, if the t-statistic obtained in our estimation is positive, then we cannot reject the null hypothesis (i.e., we cannot reject that the process is non-stationary).

In this kind of test, we assume that there is no serial correlation in the error term u_t. However, the process described by equation 7 may be non-stationary, even if there is serial correlation in u_t. As an extension of the methodology, we can now allow for serial correlation, and perform a unit roots analysis, using the so-called *augmented Dickey-Fuller test* (ADF). For that purpose, we can expand the model by adding the lagged dependent variable to the right-hand side of the equation, as follows:

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \sum_{j=1}^{m} \lambda_j \Delta Y_{t-j} + u_t \qquad (8)$$

where λ_j represent the ***m*** parameters obtained in the regression analysis between the dependent variable ΔY_t and the same dependent variable with a lag of ***j*** periods (i.e., ΔY_{t-j}). For example, for annual data, if we consider two lags, we would have the following expression:

$$\Delta Y_t = \alpha + \beta Y_{t-1} + \lambda_1 \Delta Y_{t-1} + \lambda_2 \Delta Y_{t-2} + u_t \qquad (9)$$

where we have added, on the right-hand side of the equation, two terms that include the dependent variable with a lag of one year and two years (ΔY_{t-1} and ΔY_{t-2}, respectively). The number of lags considered in the analysis depends on the decision of the analyst and the kind of problem that is being analyzed.

The regression analysis to determine the parameters in equation 8 is made using OLS. Then, the t-statistic obtained for the parameter $\boldsymbol{\beta}$ is 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, then we cannot reject that $\beta = 0$ and we cannot reject that the process is non-stationary after detrending.

RESULTS OBTAINED FOR SPANISH TOLL HIGHWAYS

In this section, the methodology described above is applied, in both versions (the Dickey-Fuller and the Augmented Dickey-Fuller tests), for traffic series in Spanish toll highways. As a starting point, we use the data collected by the public authority (Delegación del Gobierno en las Sociedades Concesionarias de Autopistas de Peaje, 2008) which is in charge of the supervision of national toll highways. These highways have an average length of 134 km, and all of them are managed by private companies under concession contracts. These private companies are obliged to provide the relevant data to the mentioned public authority, and these data are published, and available for researchers or for any person interested in this matter.

We use the Annual Average Daily Traffic (AADT) for our research. By using annual data, we avoid the problem of seasonality in traffic volumes. The collected data are included in Appendix 1 in this paper.

In order to perform the DF test, we call $Y_t = ln\ (\theta_t)$, where θ_t is the volume of traffic, in terms of AADT. Therefore, we use equation 7, where $\Delta Y_t = ln\ (\theta_t / \theta_{t-1})$. We have applied a regression analysis, using OLS to obtain the estimation of the parameter $\boldsymbol{\beta}$ and the t-statistic for that estimation for each highway. The results for the relevant t-statistics are included in the third column of Table 2.

Table 2 - Results of unit roots tests for traffic series

Name of Highway	Period of analysis	DF test t-statistic	ADF test (one lag) t-statistic	ADF test (two lags) t-statistic
Villalba-Adanero	1974-2007	0,6692	0,7843	0,8895
Zaragoza-Mediterráneo	1976-2007	-1,5419	-0,8040	-1,2278
Sevilla-Cádiz	1974-2007	1,1856	0,3834	0,0468
Montmeló-La Junquera	1974-2007	0,6792	-0,5228	-0,2624
Barcelona-Tarragona	1974-2007	-1,4503	-1,7321	-1,2479
Montmeló-Papiol	1978-2007	-0,7704	-1,2694	-1,4408
Bilbao-Zaragoza	1978-2007	0,8923	-0,4104	-0,1090
Burgos-Armiñón	1978-2007	-2,0183	-0,6922	-0,3636
León-Campomanes	1983-2007	0,2918	-1,3950	-1,1522
Tarragona-Valencia	1974-2007	-0,3193	-0,9579	-0,8513
Valencia-Alicante	1976-2007	-1,3442	-0,8213	-0,1557

We can compare these results with the critical values in Table 1. As the period of analysis is around thirty years in most cases, we can take the critical values for a sample size equal to 25 in Table 1. As we can observe, for significance levels of 5% and 10%, we cannot reject the null hypothesis (i.e., $\beta = 0$) for any of the highways that are analyzed. This means that, according to the DF test, we cannot reject the hypothesis that traffic in Spanish toll highways follows a GBM process.

We have also performed the ADF test, now using equation 8, where again $\Delta Y_t = ln\ (\theta_t / \theta_{t-1})$. We have taken one lag and two lags for the analysis, which is considered to be enough, when we observe the results obtained.

With one lag, the regression analysis is applied using the following expression:

$$\Delta(\ln \theta_t) = \alpha + \beta(\ln \theta_{t-1}) + \lambda_1 \Delta(\ln \theta_{t-1}) + u_t \qquad (10)$$

Here we estimate the parameter $\boldsymbol{\beta}$ and calculate its t-statistic.

With two lags, the relevant expression is the following:

$$\Delta(\ln \theta_t) = \alpha + \beta(\ln \theta_{t-1}) + \lambda_1 \Delta(\ln \theta_{t-1}) + \lambda_2 \Delta(\ln \theta_{t-2}) + u_t \qquad (11)$$

Again, we are interested in the estimation of the parameter $\boldsymbol{\beta}$.

The relevant t-statistics for each highway are included in the fourth and fifth columns in Table 2. As we can see, if we compare again with the critical values in Table 1, we cannot reject the null hypothesis for any of the highways. Therefore, we cannot reject that traffic follows a GBM process. On the other hand, there is not a clear pattern in the values of the t-statistic with one lag and with two lags. For some highways, the t-statistic is nearer the critical value with two lags than with one lag, and in other cases it is the other way round.

LIMITATIONS OF THE ANALYSIS AND APPLICATION OF THE RESULTS

According to the results obtained in the research described in this paper, we cannot reject the GBM hypothesis for traffic volume. However, we should be aware about the limitations of the analysis. This result is only a weak evidence in favor of the hypothesis that traffic actually follows a GBM. In fact, the results could be different for longer periods of analysis, as the results obtained by Pyndick and Rubinfeld (1998) show for the case of commodity prices. Unfortunately, the availability of longer traffic series is not usual.

Nevertheless, the results are robust, in the sense that we have applied the relevant tests to all the national toll highways in Spain, and we could not reject the hypothesis in any of them. We would be tempted to generalize the results, since there are various types of highways in the sample used: some of them are coastal highways (with a clear touristic character), some others are interurban highways and, finally, other highways have some of the features of metropolitan transportation networks.

Another limitation of the analysis is the assumption of a constant volatility of traffic. For the estimation of this volatility, we have used historic data in Spanish toll highways. A simple procedure to calculate the traffic volatility is the following:

Let us suppose that we have a traffic series for a certain highway: θ_1, θ_2, θ_3, … θ_n, where θ_i is the traffic volume in year i. Then, we define the following variable: $x_i = \Delta \ln(\theta_i) = \ln(\theta_i / \theta_{i-1})$, and we obtain $\bar{x}$ as the mean of $x_1, x_2 \ldots . x_n$. Then the volatility of traffic, defined as the standard deviation of the sample $x_1, x_2 \ldots . x_n$, would be the following:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{i=n} (x_i - \bar{x})^2} \qquad (12)$$

Using this definition, we have obtained the volatility for traffic in each toll highway in Spain, starting from data contained in Appendix 1. We have assumed that the volatility in each highway remains constant, but in fact it may change over time. However, we have observed that traffic volatility in toll highways is greater during the first years of the concession, becoming smaller and stabilized afterwards. This means that, if we have time series long enough (say twenty years) we can assume the hypothesis of a constant volatility in future. In our case, we have obtained that the volatility of traffic in Spanish toll highways (for annual data) tends towards an average value close to 0,075.

The hypothesis of the Geometric Brownian Motion given by equation 1 can be applied for the valuation of toll highways concessions. In this kind of concession, both the forecast of future traffic and the measure of the risks involved are essential for the appraisal of the business.

The calculation of the value of the volatility of traffic (probably the most important source of uncertainty in a toll highway) allows for using the model to build a confidence interval for the traffic forecast.

Besides, the terms of reference in toll highway concessions (and the concession contracts) often contain certain clauses that allow for a degree of operational flexibility in the management of the business. The valuation of this kind of clauses in contracts can be carried out using a real options approach, a methodology based on the development of the theory of financial options. Under this approach, traffic volume on the highway (for which a GBM process is assumed) is used as the underlying asset in an option contract. Options that are embedded in the concession agreement are thus calculated as a derivative of the traffic volume. This means that traffic is treated as the source of uncertainty that determines the value of the options.

The full description of this methodology is beyond the scope of this paper, but some of the options that usually appear in concession contracts may be quoted: public participation loans, shadow tolls, minimum traffic guarantees (traffic floors), maximum traffic limitations (traffic caps), extension of the concession, anticipated reversion, granting of public subsidies, etc. These mechanisms reduce the variability of the project cash-flows, and allow for more flexibility and a better management of the concession based on the contingency of future events.

The possible exercise of this series of rights represents an added value for the project which is not captured by the traditional procedures of valuation. The habitual practice of calculating the net present value (NPV) of the project by means of the discount of cash flows, leads to erroneous results when the project incorporates a certain degree of flexibility.

Therefore, the theory of real options is an alternative tool for the correct valuation of toll highway concessions, under the hypothesis that the variations of traffic volume follow a GBM like the one described in former equation 1.

CONCLUSIONS

The main result of our research is that we cannot reject the hypothesis that traffic follows a generalized Wiener process (or so-called Geometric Brownian Motion) in Spanish toll highways. In other words, the evidence that we have found leads to the conclusion that *we cannot reject the non-stationarity hypothesis* for traffic, but we have to keep in mind that this is only a weak evidence in favor of the hypothesis that traffic *actually* follows a non-stationary process.

The hypothesis of a Geometric Brownian Motion for traffic can be applied for the valuation of toll highway concessions. Often, concession contracts in highways include some kind of clauses (for example, a minimum traffic guarantee) that allow for a better management of the business risks. The value of these clauses may be important and should be added to the total value of the concession. This kind of valuation can be performed using a methodology based on a real options approach. The results of our research allow for the application of this methodology under the assumption that the evolution of traffic volume follows a Geometric Brownian Motion.

REFERENCES

Black, F. and Scholes, M. (1973): “The pricing of options and corporate liabilities”. Journal of Political Economy, vol. 81, pp. 637-654.

Delegación del Gobierno en las Sociedades Concesionarias de Autopistas de Peaje (2008): Informe 2007 sobre el sector de autopistas de peaje en España. Ministerio de Fomento, Madrid.

Dickey, David A. and Fuller, Wayne A. (1979): “Distribution of the Estimators for Autoregressive Time-Series with a Unit Root”. Journal of the American Statistical Association, vol. 74, pp. 427-431.

Dickey, David A. and Fuller, Wayne A. (1981): “Likelihood Ratio Statistics for Autorregressive Time Series with a Unit Root”. Econometrica, vol. 49, pp. 1057-1072.

Dixit, Avinash K., Pindyck, Robert S. (1994): Investment under uncertainty, Princeton University Press. Princeton, New Jersey.

Hull, John C., 2006: Options, Futures and Other Derivatives, Pearson Prentice Hall. 6th edition. Upper Saddle River, New Jersey.

Itô, K., 1951: “On Stochastic Differential Equations”. Memoirs of the American Mathematical Society, 4 (1951), pp. 1-51.

Merton, R.C. (1973): “Theory of rational option pricing”. Bell Journal of Economics and Management Science, vol. 4, pp. 141-183.

McKinnon, J. (1990): Critical values for cointegration tests. Discussion paper 90.4, University of California, San Diego.

Pindyck, Robert S. and Rubinfeld, Daniel L. (1998): Econometric Models and Economic Forecasts. McGraw-Hill, 4th edition, Boston.

Zhao, T., Sundararajan, S.K., Tsen, C. (2004): “Highway Development Decision-Making under Uncertainty: a Real Options Approach”. Journal of Infrastructure Systems. Vol. 10, No. 1, pp. 23-32.

APPENDIX 1. TRAFFIC DATA

Annual Average Daily Traffic (AADT) in Spanish Toll Highways (I)

	Villalba-Adanero	Zaragoza-Mediterráneo	Sevilla-Cádiz	Montmeló-La Junquera
1974	7.258		3.171	14.728
1975	7.817		3.382	13.354
1976	8.168	5.276	3.017	13.002
1977	6.690	6.179	3.039	13.925
1978	7.796	6.439	3.470	15.823
1979	8.455	7.001	3.681	15.859
1980	8.326	7.053	3.774	15.026
1981	8.380	6.920	3.999	15.557
1982	8.355	6.761	3.929	15.948
1983	8.283	6.607	3.629	15.934
1984	8.452	6.489	3.417	16.478
1985	8.810	6.659	3.632	17.099
1986	9.478	7.181	3.959	18.892
1987	10.360	8.119	4.525	21.282
1988	11.420	9.387	5.282	23.671
1989	12.929	11.423	6.350	26.296
1990	14.005	12.127	6.835	26.660
1991	15.610	12.327	7.791	27.802
1992	16.415	12.174	9.214	28.488
1993	16.504	11.425	8.005	28.124
1994	16.628	10.958	7.978	28.554
1995	17.358	11.309	7.648	28.509
1996	17.866	11.027	7.434	27.076
1997	18.687	11.423	7.828	29.021
1998	20.715	12.377	10.101	30.717
1999	22.918	13.350	11.825	33.815
2000	24.325	14.870	13.300	35.955
2001	25.482	15.206	15.218	37.901
2002	27.238	15.594	16.534	40.464
2003	28.662	15.464	17.897	41.756
2004	30.301	15.350	19.642	43.324
2005	30.770	14.744	21.859	44.918
2006	32.998	15.273	24.244	47.122
2007	34.414	15.541	24.951	49.180

Annual Average Daily Traffic (AADT) in Spanish Toll Highways (II)

	Barcelona-Tarragona	Montmeló-Papiol	Bilbao-Zaragoza	Burgos-Armiñón
1974	15.377			
1975	15.367			
1976	16.630			
1977	19.760			
1978	22.811	9.389	4.689	2.479
1979	23.659	6.875	4.169	3.604
1980	24.565	7.480	4.606	4.060
1981	23.575	6.470	4.681	5.622
1982	23.613	6.723	4.754	4.966
1983	23.166	6.861	4.374	4.611
1984	23.597	6.944	4.281	4.970
1985	24.857	7.352	4.275	5.142
1986	27.154	27.404	4.433	5.487
1987	30.793	31.558	4.874	5.994
1988	34.963	42.998	5.617	6.832
1989	39.624	51.004	6.494	7.777
1990	40.618	52.226	6.870	8.294
1991	42.080	54.489	7.118	8.954
1992	41.379	49.997	7.052	9.403
1993	40.152	45.884	6.956	9.680
1994	41.123	46.960	6.930	10.172
1995	43.270	48.724	7.013	11.026
1996	43.530	52.453	7.038	11.430
1997	45.677	58.635	7.343	12.198
1998	47.799	63.220	8.082	13.696
1999	47.089	70.219	9.002	15.161
2000	51.278	83.935	10.623	16.605
2001	53.721	90.218	11.742	18.062
2002	55.994	92.636	12.196	19.348
2003	57.782	95.712	12.844	20.101
2004	59.053	99.460	13.503	21.072
2005	60.342	111.353	13.542	21.206
2006	63.683	115.607	14.177	22.209
2007	66.217	118.519	14.712	23.937

Annual Average Daily Traffic (AADT) in Spanish Toll Highways (III)

	León-Campomanes	Tarragona-Valencia	Valencia-Alicante
1974		5.603	
1975		5.776	
1976		6.002	3.563
1977		6.870	4.148
1978		7.524	5.183
1979		7.828	5.874
1980		7.773	6.059
1981		7.590	6.258
1982		7.455	6.147
1983	2.494	7.233	6.071
1984	2.049	7.178	6.124
1985	2.141	7.596	6.933
1986	2.275	8.514	7.240
1987	2.445	9.707	8.316
1988	2.768	10.873	9.376
1989	3.233	12.336	10.563
1990	3.661	12.501	12.027
1991	4.254	13.043	12.663
1992	4.256	12.894	12.595
1993	4.199	12.336	12.085
1994	4.583	12.469	12.301
1995	4.680	12.907	12.313
1996	4.718	13.070	12.423
1997	4.995	14.186	13.207
1998	5.659	16.692	16.271
1999	6.320	19.092	18.987
2000	6.642	20.453	21.225
2001	7.433	22.004	23.409
2002	7.679	22.796	24.968
2003	8.048	23.396	26.640
2004	8.736	23.932	27.302
2005	9.006	23.482	28.180
2006	9.683	25.215	29.207
2007	10.288	25.110	29.411

.