Modelling International Monthly Tourist in Spain*

JUNCAL CUÑADO  
Departamento de Economía y Métodos Cuantitativos, UNIVERSIDAD DE NAVARRA, ESPAÑA. E-mail: jcunado@unav.es  

LUIS ALBERIKO GIL-ALANA  
Departamento de Economía y Métodos Cuantitativos, UNIVERSIDAD DE NAVARRA, ESPAÑA. E-mail: alana@unav.es  

FERNANDO PEREZ DE GRACIA  
Departamento de Economía y Métodos Cuantitativos, UNIVERSIDAD DE NAVARRA, ESPAÑA. E-mail: fgracia@unav.es

ABSTRACT

This article investigates the degree of persistence in the international monthly tourist time series in Spain using long memory (fractional integration) techniques. Our findings can be summarized as follows. The two standard hypotheses of integer degrees of differentiation, i.e., the I(0) and the I(1) behaviour, are clearly rejected. The series is found to be I(d) with a value of d in the interval (0.421, 0.780) thus implying long memory behaviour and mean reversion behaviour. However, if a structural break is considered, it takes place at May 2007, and then, the two subsamples present orders of integration which are above 1 and thus rejecting the mean reverting hypothesis.

Keywords: Persistence, Tourists, Fractional Integration.

Modelización de llegadas mensuales de turistas a España

RESUMEN

Este trabajo investiga el grado de persistencia en las llegadas internacionales de turistas a España empleando técnicas de integración fraccional. Los resultados que se obtienen son los siguientes. Los resultados sugieren que las hipótesis de I(0) e I(1) se rechazan y que la serie de llegada de turistas a España se puede modelizar como una serie I(d) donde d toma valores en el intervalo (0.421, 0.780) implicando un proceso de memoria larga y con reversión a la media. Sin embargo, teniendo en cuenta la existencia de un cambio estructural, este ocurre en Mayo del 2007, y las dos submuestras son entonces integradas fraccionadamente con un parámetro de integración superior a 1 en ambas submuestras, y rechazando por tanto la hipótesis de reversión a la media.

Palabras clave: Persistencia, turistas, integración fraccional.

JEL Classification: C32, E62

* The authors gratefully acknowledge financial support provided by the Ministerio de Ciencia y Tecnología (ECO2008-02458-E). We also thank the Editor and two anonymous referees for helpful comments.

Artículo recibido en mayo de 2011 y aceptado en noviembre de 2011  
Artículo disponible en versión electrónica en la página www.revista-eea.net, ref. ə-29316

ISSN 1697-5731 (online) – ISSN 1133-3197 (print)
1. INTRODUCTION

Tourism is one of the most important sectors in the Spanish economy. Recent data on international tourism shows that Spain received 5.5 millions of tourists in January and February 2011, which represents an increased of 4.5% with respect to last year (see, Ministerio de Industria, Turismo y Comercio, 2011). Recent data also show that the contribution of tourism to Gross Domestic Product is almost 10% in 2009 (see, Instituto Nacional de Estadística, 2010). Furthermore, according to the World Tourism Organization (OMT, 2010), Spain ranks fourth in the World among countries, with 53 million of international tourists (78 million in France, 60 million in US and 55 million in China).

The economic literature on modelling and forecasting arrivals and number of tourists (national and international) is extensive. The literature of tourism modelling and forecasting began as early as the 1960s (see, for example, Gray, 1966 and Laber, 1969). For example, we find approaches based on error correction models and cointegration techniques (see for example Syriopoulos, 1995; Kulendran, 1996; Kulendran and King, 1997; Seddighi and Shearing, 1997; Kim and Song, 1998 and Vogt and Wittayakorn, 1998 among others), while other authors use pure time series analytical models (see García-Ferrer and Queralt, 1997; Chu, 1998; Kim, 1999, Lim and McAleer, 2001, 2002; Goh and Law, 2002; Gustavsson and Nordström, 2001 and Brännäs et al., 2002).

Recently, and following the development on model building and time series prediction, tourism (and arrivals) modelling and forecasting has experienced considerable methodological changes in recent years. Thus, for example, Goh and Law (2002) use different univariate models (MA, smoothing, ARIMA and seasonal ARIMA –SARIMA- models) to forecast monthly arrivals of ten of the highest tourist-generating countries for Hong Kong and evaluate its accuracy, obtaining the best fit corresponding to the two latter methods. According to these models, they obtain that all the ten series have a seasonal unit root at the zero and seasonal frequencies. Similarly, Lim and McAleer (2002) test for the presence of unit roots and seasonal unit roots for tourist arrivals to Australia from Hong Kong, Malaysia and Singapore. They use Fuller (1976) and Dickey and Fuller (1979) tests for the non-seasonal root, and Hyllerberg et al. (1990) (HEGY, 1990) for the seasonal ones, obtaining that quarterly tourist arrivals are each integrated of order 1 at the zero and the semi-annual frequencies. Therefore, both studies perform seasonal differencing to render the series stationary. On the other hand, Gustavsson and Nordström (2001) indicate that ARMA models, built conditionally on the outcome from seasonal unit root tests, do not produce more accurate forecasts than models that a priori impose unit roots at zero and all seasonal frequencies, when analysing tourism in Sweden.

In this paper we analyze the persistence in the total number of international Spanish arrivals by means of fractional integration over the recent period.
2000:M1 to 2011:M2. In other words, instead of imposing an integer degree of differentiation, i.e. 0 in case of stationarity, or 1 with nonstationarity (unit roots), we allow the order of integration to be fractional. Fractional integration has been applied in tourist series in a wide number of studies (see, for example, Cunado et al, 2004, 2005, 2008a,b; Gil-Alana et al., 2008 and recently, Gil-Alana, 2009, 2010 among others). For example, Cunado et al. (2005) focused on the nature of seasonality in Spanish tourism time series using deterministic and/or stochastic (with integer and fractional differentiation) seasonality. Their results showed that the Spanish tourism time series (i.e., the total foreign and domestic number of tourist; the number of domestic tourists, the number of nights spent in hotel accommodation by tourists and the number of nights spent in hotel accommodation by domestic tourists) can be described in terms of seasonally fractionally integrated models. In another recent study, Cunado et al. (2008a) modelled the total number of international monthly arrivals in the US using fractional integration with a structural break that is endogenously determined by the model. Their results show that the break takes place at September 2001. Disaggregating the data according to the location of origin for the arrivals, the break at September 2001 is also found in most cases, the only exceptions being Canada (February, 1998) and Middle East (September, 2000). Furthermore, they find that the orders of integration of all series are strictly smaller than 1 implying mean reversion. Alternatively, Cunado et al. (2008b) investigated persistence in international monthly arrivals to the Canary Islands using a model based on fractional integration and seasonal autoregressions. Their evidence support that the aggregate series corresponding to the total number of international arrivals in the Canary Islands can be approached as I(d) process with d slightly above 0.5.

The remainder of the paper is organized as follows. Section 2 describes the statistical approach employed in the paper. In Section 3, we apply the proposed methodology to the monthly Spanish data and report the empirical results. Finally, Section 4 contains some concluding comments.

2. THE STATISTICAL MODEL

In this section we describe a simple procedure that allows us to estimate deterministic trends and possibly fractional orders of integration. We suppose that \( y_t \) is the observed time series, generated by the model:

\[
y_t = \alpha + \beta t + \chi_t; \quad (1 - L)^d \chi_t = \epsilon_t, \quad (1)
\]

where the \( \alpha \) and the \( \beta \) are the coefficients corresponding to the intercept and the linear trend; \( d \) may be real values and \( \epsilon_t \) is I(0). The second expression in (1) is the standard case of \( I(d) \) models of the form:
\[(1 - L)^d x_t = u_t, \quad t = 0, \pm 1, \ldots, \quad (2)\]

with \(d > 0\), where \(L\) is the lag-operator \((Lx_t = x_{t-1})\) and \(u_t\) is \(I(0)\). Note that the polynomial in the left-hand-side in (1) can be expressed in terms of its Binomial expansion, such that, for all real \(d\),

\[
(1 - L)^d = \sum_{j=0}^{\infty} \psi_j L^j = \sum_{j=0}^{\infty} \frac{d^j}{j!}(-1)^j L^j = 1 - dL + \frac{d(d-1)}{2}L^2 - \ldots,
\]

and (2) can be written as:

\[
x_t = dx_{t-1} - \frac{d(d-1)}{2} x_{t-2} + \ldots + u_t.
\]

If \(d\) is an integer value, \(x_t\) will be a function of a finite number of past observations, while if \(d\) is not an integer, \(x_t\) depends strongly upon values of the time series far away in the past. (See, e.g., Granger and Ding, 1996; Dueker and Startz, 1998). Moreover, the higher the \(d\) is, the higher the level of association will be between the observations. Thus, \(d\) is an indicator of the degree of persistence of the series.\(^1\)

The parameter \(d\) also plays a crucial role from the statistical viewpoint. Thus, if \(d < 0\), the series is said to display “anti-persistence” (Mandelbrot, 1977), or to exhibit intermediate memory or long-range negative dependence. In such a case, the time series reverses itself more often than a random series would. If \(d = 0\), \(x_t\) is stationary \(I(0)\) and is commonly denoted as “short memory”; on the contrary, if \(d > 0\), \(x_t\) is said to be “long memory”, so-named because of the strong degree of association between observations in the far distant past, and if \(0 < d < 0.5\), the series is still covariance stationary; however, if \(d \geq 0.5\), the series is no longer stationary and as \(d\) increases beyond 0.5 and through 1 the series is becoming “more nonstationary” in the sense, for example, that the partial sums increase in magnitude with \(d\) being non-summable.

In summary, if \(d < 0.5\), \(x_t\) is covariance stationary and mean-reverting, with the effect of the shocks dying away in the long run. If \(d \in \{0.5, 1\}\), \(x_t\) is no longer covariance stationary but is still mean reverting, while \(d \geq 1\) implies nonstationarity and non-mean-reversion. In particular, a variable having a unit-root supports the view that any shock will have a permanent effect, so a policy action will be required to bring the variable back to its original long-term behaviour. On the other hand, if \(d\) is smaller than 1, fluctuations will be transitory and therefore, there exist less need for the policy action, since the series will in any case return to its trend sometime in the future.

\(^1\) For recent surveys of fractional integration see the papers of Robinson (2003), Doukhan et al. (2003) and more recently, Gil-Alana and Hualde (2009).
3. EMPIRICAL RESULTS

The time series data analysed in this section correspond to the monthly observations for the international tourism in Spain, for the recent period 2000:M1-2011:M2.² All this data have been obtained from Movimientos Turísticos en Fronteras (FRONTUR).³ A preliminary analysis of the data shows the possibility of a structural break in the series, so that, in this Section we carry out the estimation procedure using first the whole time period, and in a second sub-section, allowing for a structural endogenous break.

3.1. Empirical results

Figure 1 contains plots of the original series (total tourists in Spain) along with the correlograms and periodogram. We see that the original series presents a clear seasonal pattern, which is also viewed across the correlogram (with significant values at the seasonal lags) and through the periodogram (with large peaks at the monthly frequencies). In order to deal with the seasonal problem, seasonal AR models will be used to describe the short run dynamics of the series.

\[\text{Figure 1}\]

Original time series with its corresponding correlogram and periodogram (January 2000- February 2011)

\[\text{Time series plot}\]

² See, for example, Cabrè and Iranzo (2007) for a recent study on tourism in Spain.

³ See, Instituto de Estudios Turísticos, Ministerio de Industria, Turismo y Comercio (Gobierno de España).

Estudios de Economía Aplicada, 2011: 723-736 • Vol 29-3
Correlogram of the original data

Periodogram of the original data

The large sample standard error under the null hypothesis of no autocorrelation is $1/\sqrt{T}$ or roughly 0.089. The periodogram was computed based on the discrete Fourier frequencies $\lambda_j = 2\pi j/T$.

Source: Own elaboration.

We consider a model of form:

$$y_t = \alpha + \beta t + x_t; \quad (1-L)^d x_t = u_t; \quad t = 1, 2, ..., \quad (3)$$

where $d$ above represents the order of integration of the series, which may be 0 (stationary I(0)), 1 (unit roots, I(1)) or any other real value. For $u_t$, we choose a
white noise process, along with non-seasonal and seasonal (monthly) AR(1) disturbances. Higher AR orders were also implemented and the results did not substantially differ from those employed here. We estimate the parameters in (3) by maximum likelihood for the two cases of an intercept (i.e., $\alpha$ unknown and $\beta = 0$ a priori) and with an intercept and a linear trend ($\alpha$ and $\beta$ unknown) for the total number of arrivals in Spain. The results for the estimates of $d$ are displayed in Table 1. In the table we report the estimates of $d$ along with the 95% confidence intervals of the non-rejection values of $d$ using Robinson’s (1994) parametric approach. The intervals were constructed testing different values of $d$ using a grid of 0.01-values. The main results can be summarized as follows.

### Table 1
Estimates of $d$ and 95% confidence interval using Robinson’s (1994) approach

<table>
<thead>
<tr>
<th></th>
<th>No regressors</th>
<th>An intercept</th>
<th>A linear time trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Noise</td>
<td>1.269* (1.041, 1.512)</td>
<td>1.274* (1.039, 1.519)</td>
<td>1.273* (1.039, 1.518)</td>
</tr>
<tr>
<td>AR (1)</td>
<td>0.446 (0.364, 0.523)</td>
<td>0.445 (0.366, 0.526)</td>
<td>0.453 (0.375, 0.532)</td>
</tr>
<tr>
<td>Monthly AR (1)</td>
<td>0.799 (0.689, 0.943)</td>
<td>0.563 (0.410, 0.780)</td>
<td>0.586 (0.421, 0.786)</td>
</tr>
</tbody>
</table>

*Note: $d$ estimates are in bold and 95% confidence intervals are in brackets. * significant values of $d$ which are statistically different from 1.*

*Source: Own elaboration.*

We observe that if the disturbance term is a white noise process, the unit root null hypothesis is rejected in favour of orders of integration above 1 for the three cases of no regressors, an intercept, and an intercept with a linear trend. However, modelling the error term with autocorrelated errors, the values are strictly below 1, being about 0.4 with non-seasonal $u_t$, and slightly higher with seasonal AR(1) errors.

### Table 2
Estimates in the model with seasonal (monthly) AR(1) disturbances

<table>
<thead>
<tr>
<th></th>
<th>$d$</th>
<th>Intercept</th>
<th>Time trend</th>
<th>AR coeff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>With a linear time trend</td>
<td>0.586 (0.421, 0.786)</td>
<td>2.960E06 (3.239)</td>
<td>8087.121 (0.487)</td>
<td>0.971</td>
</tr>
<tr>
<td>With an intercept</td>
<td>0.563 (0.410, 0.780)</td>
<td>3.234E06 (3-971)</td>
<td>-----</td>
<td>0.972</td>
</tr>
</tbody>
</table>

*Source: Own elaboration.*
Performing several diagnostic tests on the residuals, the results indicate that the model with monthly AR(1) disturbances seems to be the most adequate specification, and given that the time trend coefficient is found to be statistically insignificant, the selected model for the Spanish international monthly arrival over the period 2000:M1 – 2011:M2 is:

\[
y_t = 3-234E06 + x_t; \quad (1 - L)^{0.563}x_t = u_t; \quad u_t = 0.972u_{t-12} + \varepsilon_t.
\]

where the first parenthesis refers to the t-value of the intercept, and the second to the 95% confidence band for the fractional differencing parameter.

In a related paper, Cunado et al. (2008a) obtained very similar results for the international monthly arrivals in the US. They found that the values of d were in all cases constrained between 0 and 1 and thus implying fractional integration. In case of the total number of arrivals, the values of d were 0.58 with an intercept, and 0.60 with a linear trend. Cunado et al. (2008b) also obtained similar results for the international monthly arrivals to the Canary Islands with d slightly above 0.5.

In summary, the estimated results support that the two standard hypotheses of I(0) and I(1) behaviour are clearly rejected. The process is highly persistent, with d slightly above the nonstationarity limit and with a seasonal AR(1) coefficient very close to 1.

In Figure 2 we display the first 120 values of the impulse responses based on the selected model. It is observed that the values decrease very slowly with a very strong seasonal pattern. In fact, the combination of the two parameters, the fractional differencing and the seasonal AR one, makes the impression in the figure that the responses increase over time, though they disappear though in the very long run.

**Figure 2**

Impulse responses and 95% confidence band

![Impulse responses and 95% confidence band](image)
3.2. Estimation with a structural break

Within the context of fractional integration, we also consider the possibility of structural breaks. Note that fractional integration and structural breaks are closely related issues. For example, Bhattacharya et al. (1983), Teverovsky and Taqqu (1997), Diebold and Inoue (2001), Granger and Hyung (2004) and Ohanissian et al. (2008) among many others show that fractional integration may be a spurious phenomenon caused by the existence of breaks in short-memory I(0) contexts. In the same line, Kuan and Hsu (1998), Wright (1998) and Krämer and Sibbertsen (2002) showed that evidence of structural change might be spurious since most commonly employed tests for breaks are biased towards an over-rejection of the null of no change when the process exhibits long memory. Here, we employ a recent procedure developed by Gil-Alana (2008) that is based on minimizing the residuals sum squares across different subsamples. For simplicity, we describe here the case of a single break and consider a model of the form:

\[ y_t = \beta_1^T z_t + x_t, \quad (1 - L)^{d_1} x_t = u_t, \quad t = 1, \ldots, T_b \]  

(4)  

and

\[ y_t = \beta_2^T z_t + x_t, \quad (1 - L)^{d_2} x_t = u_t, \quad t = T_b + 1, \ldots, T, \]  

(5)  

where the \( \beta \)'s are the coefficients corresponding to the deterministic terms; \( d_1 \) and \( d_2 \) may be real values; \( u_t \) is I(0); and \( T_b \) is the time of a break that is supposed to be unknown. This method is based on minimizing the residuals sum squares from a grid of values of the fractional differencing parameters at different break dates.

We focus on the case of a single break based on the sample size examined. The inclusion of more than one break would result in subsamples which would be too small to justify fractional integration (see Table 3).

\begin{table}[h]  
\centering  
\begin{tabular}{|l|l|l|l|l|l|}  
\hline  
& & & & & \\
1\textsuperscript{st} sub-sample & 2\textsuperscript{nd} sub-sample & & & & \\
\hline  
d & Intercept & AR coef. & d & Intercept & AR coef. \\
\hline  
1.257 & 1823835.8 & 0.920 & 1.348 & 7826337.0 & 0.891 \\
(1.012,1.501) & (1.973) & & (1.132,1.618) & (9.305) & \\
\hline  
\end{tabular}  
\caption{Estimates of d with a structural break (May 2007)}  
\end{table}  

\textit{Source:} Own elaboration.

The break takes place at May 2007, which is consistent with the plot of the data in Figure 1. Surprisingly, the estimated values of \( d \) are found to be higher than 1 in the two subsamples. It is 1.257 in the first subsample and 1.348 in the
second one. However, the fact that the confidence intervals overlap indicate that we cannot reject the null of equal orders of integration. Also, the seasonal AR coefficient is high and close to 1 in the two subsamples. So, in spite of the break, the two sub-series are highly persistent, nonstationary and non-mean reverting. This is an interesting result in the sense that it shows that fractional integration may be a mechanism generated by the existence of breaks in the data. Thus, if no breaks are considered the series seems to be long memory and mean reverting, with an order of integration strictly smaller than 1. However, testing for a break, this seems to take place at May 2007, and the orders of integration are strictly higher than 1 in the two subsamples, and thus rejecting mean reversion in its behavior.

What can we conclude from our results? First, the procedure of Gil-Alana (2008) strongly supports the existence of a single break in the data at May 2007. Moreover, performing some diagnostic tests on the estimated residuals from the two subsamples (tests of no additional serial correlation, Durbin, 1970; Godfrey, 1978a,b and homoscedasticity, Koenker, 1981) the results pass the diagnostics in all cases for the error terms. Also, based on the independence of the two subsamples when using this approach, we also carried out the estimation of d using Robinson’s (1994) approach and the results were completely in line with those reported in Table 3. This method is valid for any real value of d, and therefore, it does not require a preliminary first differentiation as is the case in other procedures of fractional integration or in the standard cases of integer orders of integration where the estimation must be carried out based on both the original and the first differenced data.

4. CONCLUSIONS

In this paper we investigate the degree of persistence in the international tourist time series, using Spanish monthly data for the period January 2000 - February 2011. We approach the international monthly arrivals using a fractionally integrated model with seasonal AR disturbances. The results indicate that the series is I(d) with d constrained between 0 and 1 and thus support the fractionally integrated approach. The estimated value of d is about 0.563 which is jointly estimated with a seasonal AR coefficient very close to the unit circle (0.972). That means that the time series is highly persistent though mean reverting and with a strong seasonal pattern. Similar results are obtained in the literature by Cunado et al. (2008a,b). However, a very different picture emerges when a structural break is taken into account. Here we found a break at May 2007, and the results indicate that the estimated value of d is higher than 1 before and after the break. We also find that the seasonal AR coefficient is close to 1 in both subsamples suggesting than the international tourist series in the Spanish case are highly persistent and non-mean-reverting. Performing some
diagnostic tests on the estimated residuals in the two subsamples the results strongly support the hypothesis of a break, implying that the previous finding of long memory with mean reversion might be a spurious phenomenon caused by the existence of the structural break.

The research described in this paper can be extended in several directions. For example, it could be to apply to international arrivals time series data at a disaggregated regional level. Furthermore, a follow step in this work would be to use the selected model reported in the paper to predict the future behaviour of the time series.

REFERENCES


GODFREY, L.G. (1978a): "Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables" Econometrica 43, 1293-1301.

GODFREY, L.G. (1978b): "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables" Econometrica 43, 1303-1310.


