Autoregressive and volatility analysis of the international tourist expenditure in Spain

Jesús Barreal[*†], Adolfo Hernández[*‡], Gil Jannes[*§]

draft

Abstract

Tourism in Spain presents a highly seasonal component in terms of the number of tourists. However, with respect to the average tourist expenditure, the seasonal character has so far not been studied. The volatility of the tourist expenditure in Spain is also unknown. In this study, we therefore undertake to establish a model for the average daily expenditure from international tourists in Spain and its volatility. We determine autoregressive time series models (ARIMA) for Spain in general and for the Spanish Autonomous Communities most relevant for Spanish tourism, and make a first study of the variability of the variance or heteroscedasticity through GARCH models.

Keywords: ARIMA, GARCH, Time Series, Tourism Expenditure, Volatility

[*]Department of Financial and Actuarial Economics & Statistics, Universidad Complutense de Madrid
[†]jbarreal@ucm.es
[‡]adolfher@ucm.es
[§]gil.jannes@ucm.es

1 Introduction

Tourism is a key factor in Spain's socio-economic panorama. In the year 2017, Spain climbed from 3rd to 2nd place in arrivals and held on to 2nd position in receipts to become the world's second largest destination by both international tourist arrivals and receipts [UNWTO, 2018]. Travel and tourism is one of the main motors of Spain's economy, with a total contribution of roughly 15% both in terms of GDP and in terms of employment, a number which is expected to further increase in the next decade [WTTC, 2018]. However, while the number of tourist arrivals in Spain keeps increasing, total income from international tourism has lost this upward trend in the past decade [WB, 2018].

Within the large body of statistical research on tourism (see [Song and Li, 2008] and [Goh and Law, 2011] for reviews), tourism demand is usually defined in terms of number of arrivals, with tourist expenditure and overnight stays as distant second and third most popular variables. While number of arrivals and expenditure are obviously correlated, they have a different impact on tourism decision policy [Song et al., 2010] and their fluctuations have been shown to depend on different influencing factors [Sheldon, 1993, Song et al., 2010]. We will here focus explicitly on expenditure. Expenditure, although harder to measure statistically (data on expenditure are often based on the tourists' own estimates), is economically more relevant, and as pointed out explicitly for example in [Papatheodorou and Song, 2005], low-income mass tourism characterised by high arrival numbers in combination with low per capita expenditure "may prove unsustainable and lead to the eventual decline of tourism". While not all types of mass tourism have the same profile [Perez and Juaneda, 2000], Spain might be a case in point: The Balearic Islands have been studied as a prime example of the "sun and sand" tourism model with low economic benefit and high environmental cost for the destination [Aguiló et al., 2005], and similar problems might affect a large part of Spain's Mediterrean coast.

Comprehensive studies of tourist expenditure focusing specifically on Spain [González and Moral, 1995], [García-Ferrer and Queralt, 1997] are rare and require updating in the light of the fast development of statistical methods in this area and of more recent data.

One of the main motivations for studying tourism demand is in order to make accurate forecasts and adjust national and regional policies. In this light, it is clear that –apart from tourism expenditure itself– it is also important to have an understanding of its variability or volatility, especially for a country like Spain where the economic impact of tourism is strong. Indeed, volatility describes the fluctuations and therefore the economic risk associated to the tourism industry. Curiously, in spite of its obvious importance, statistical studies of the volatility of tourism demand are relatively rare in the academic literature, the first exception being [Chan et al., 2005]. Even more curiously, we are not aware of any analysis of the volatility of tourism demand in terms of expenditure. A key innovation of our contribution is therefore precisely this, a statistical analysis of the volatility of tourist expenditure as applied to Spain.

As we will discuss in detail in Section 3, we will use an ARIMA time series model to study the evolution of tourism expenditure, and a GARCH model to study its volatility. All our calculations are performed in R. We study global data for Spain, and the Autonomous Communities which play a major role in Spain's tourism industry. First, we will give a brief review of the related literature.

2 Literature review

As mentioned earlier, most studies express tourism demand in terms of arrivals. Statistical analyses of tourism demand specifically in terms of expenditure are much less frequent. Interesting examples applied to global world tourism can be found in [Papatheodorou and Song, 2005] and more recently [Martins et al., 2017].

For a general comparison of time-series and econometric methods for measuring and forecasting tourism demand based on a forecasting competition, see [Athanasopoulos et al., 2011]. This study finds that pure time series provide more accurate forecasts than econometric models with explanatory variables, but did not consider Artificial Intelligence (AI)-based techniques such as Artificial Neural Networks (ANN). Specific case-studies, e.g. for Taiwan [Chen et al., 2012], Australia [Tularam et al., 2012], Hong Kong [Song et al., 2011], Portugal [Fernandes et al., 2008] and Catalonia [Claveria and Torra, 2014], show a variety of different conclusions: some of these find traditional time series such as ARIMA to be the most adequate, even when compared to modern ANN techniques (e.g. [Claveria and Torra, 2014]); others on the contrary emphasize the promise of ANN [Fernandes et al., 2008, Palmer et al., 2006]. The 2011 review [Goh and Law, 2011] finds that most studies show AI-based methods to outperform their econometric counterparts, but points out that this conclusion should be handled with care, amongst other reasons because comparative studies usually compare advanced AI-techniques with basic econometric and time series-techniques, without taking into account recent developments in these latter fields. They end by advocating a hybrid forecasting system which combines econometric, time series and AI-based techniques. In any case, in spite of the dramatic improvement of AI-related techniques in recent years, it seems still to be true (as concluded a decade ago in [Song and Li, 2008] based on a comprehensive survey) that "there is no single model that consistently outperforms other models in all situations", and the performance of the models might depend not only on the quality and frequency of the data, but also on the length of the forecasting horizon, and even the concrete destination/origin country pair. Again, most of these method-comparing studies focus on tourist arrival numbers, without looking at expenditure, although [Goh and Law, 2011] cites some exceptions.

There are many interesting examples where time series techniques, mainly ARIMA, are applied to tourism demand, see [Song and Li, 2008] and references therein. For a more recent example, an ARIMA exercise for Macedonia was performed in [Petrevska, 2017].

The introduction of GARCH methods in the analysis of Tourism demand is usually credited to [Chan et al., 2005]. In spite of the obvious importance of analysing and predicting the volatility of tourism demand, the adoption of this method in the study of tourism has been rare. Some exceptions are [Chang et al., 2011, 2009, Coshall, 2009, Liu and Sriboonchitta, 2013, Makoni and Chikobvu, 2018, Shareef and McAleer, 2005], with a special mention for [Hoti et al.] and [Bartolomé et al., 2009] which study the volatility of tourism demand on the Spanish islands. Again, all these articles define tourism demand in terms of arrivals. We are not aware of any analysis of the volatility of tourism expenditure, and less so with respect to Spain.

In fact, recent studies of tourism demand in Spain have focused on the nature of seasonality [Cunado et al., 2005], the influence of income and prices [Garin-Munoz and Amaral, 2000], of the GDP of the country of origin [Rodriguez and Rivadulla, 2012] or of the 2008 economic crisis [Bernier et al., 2014] on demand, while the influence of the forecasting horizon on the precision of different machine learning-based forecasting methods was analysed in [Claveria et al., 2016]. All these articles use arrivals (or overnight stays in the case of [Garin-Munoz and Amaral, 2000]) as variable of interest. As rare counter-examples, [Nicolau and Más, 2005] attempts a model for the tourist expenditure in Spain in two stages (the decision to take a holiday and the actual expenditure) and [Perez and Juaneda, 2000] has studied tourism expenditure in the Balearic islands for different profiles of mass tourism. More general academic analyses of tourism expenditure in Spain are more than two decades old [González and Moral, 1995] and [García-Ferrer and Queralt, 1997].

3 Methodology & data

3.1 ARIMA

The different strategies used in the literature for the modelling of tourism demand, whether in terms of arrivals or of expenditure, can broadly be classified into three types:

- Time Series models
- Econometric models
- Artificial Intelligence models, in particular Machine Learning and Neural Networking models

Econometric models have greater explanatory power than pure Time Series, since they allow to assess the importance of various independent influencing variables (e.g. the GDP per capita of the visitor's country of origin). However, in terms of forecasting demand, econometric models are more complicated, since they require a prior estimate of these independent variables, leading to an important risk of generating large errors. Since time series are usually found to perform equally well or even better in terms of forecasting, and are in general much simpler to handle than econometric techniques, Time Series remain the standard method for predicting tourism demand.

In recent years, a lot of attention has been devoted to Artificial Intelligence models, particularly Artificial Neural Networks (ANN). These are based on (supervised or unsupervised) automated learning methods among a network of artificial "neurons", inspired by the way in which the brain works. ANN models are highly flexible and deliver spectacular results in domains such as image and speech recognition and medical diagnosis. They are usually considered very promising as forecasting tools, also. However, they also require a lot of fine-tuning in the network architecture, parameter selection, training algorithm and test data in order to provide reliable forecasts, and it is probably fair to say that they still have a certain way to go before becoming standard use in tourism demand analysis and forecasting.

In this contribution, we will therefore stick to the more traditional time-series techniques, and perform an ARIMA (AutoRegressive Integrated Moving Average) analysis of average tourist expenditure in Spain in the period 2008–2018. While autoregressive models for time series can be dated back to the 1920s (see [Tsay, 2000] and references therein), ARIMA models were systematized and popularized in the famous 1970 book by Box and Jenkins [Box et al., 2013].

Given a dependent variable of interest y_t, an ARIMA*(p,d,q)* model can be written in terms of a redefined variable z_t:

$$z_t = (1 - B)^d y_t$$

as

$$z_t = \mu + \phi_1 z_{t-1} + \cdots + \phi_p z_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}$$

where

- μ is the overall average (which can always be set to 0 through a shift of coordinate)
- ϕ and θ are (numerical) coefficients,
- p and q are the order of the AR (AutoRegressive) and MA (Moving Average) components of the model.
- ϵ_t is the error term (assumed to behave as Gaussian noise)
- the backshift operator B is defined through the operation $B^k y_t = y_{t-k}$
- d the number of differences required to make the series stationary.

Typically $d = 0$ (if the original series y_t is stationary, in which case $z_t = y_t$); $d = 1$, corresponding to a linear trend (the model is then stationary in terms of $z_t = y_t - y_{t-1}$); or $d = 2$ for a quadratic trend.

The ARIMA model is often written as

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p) z_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^p)\epsilon_t \tag{1}$$

or, more compactly:

$$\phi(B)(1 - B)^d z_t = \theta(B)\epsilon_t$$

where a shift $z_t = y_t - \mu$ is assumed in case $d = 0$, and $\phi(B)$ and $\theta(B)$ now represent the full polynomials of Eq. 1.

Note that p represents the maximum lag at which previous y_{t-p} influence y_t independently from all inbetween y_{t-i} (i.e., the highest lag with non-zero partial autocorrelation); and similarly, q represents the maximum lag at which previous error terms ϵ_{t-q} influence ϵ_t independently from all inbetween ϵ_{t-i} (i.e., the highest lag with non-zero autocorrelation).

Tourism is a typical example of a seasonal activity, and so one expects to see this seasonality reflected in the time series and in the model that best describes it. A seasonal ARIMA model can be denoted as

ARIMA*(p,d,q)*×*(P,D,Q)[S]*,

where small letters describe the overall (non-seasonal) trend and capital letters reflect the seasonal part, and S is the seasonal time span.

Using the compact notation introduced above, this model can be written formally as

$$\phi(B)\Phi_S(B^S)(1 - B)^d(1 - B^S)^D z_t = \theta(B)\Theta_S(B^S)\epsilon_t$$

where we used the index S to indicate the seasonal parts, and the seasonal polynomials are defined in a way similar to the non-seasonal polynomials defined above. Determining the ARIMA model then consists in determining which values of p, d, q, P, D, Q and S are best adapted to the concrete time series under study, and fitting the corresponding coefficients ϕ_i, θ_i and likewise Φ_i and Θ_i.

3.1.1 Dickey-Fuller test

The first step when identifying the ARIMA model that best fits the data, is to find the required value of the (non-seasonal) difference d. The reason for differencing is that the mathematical treatment within the ARIMA framework requires the time series to be stationary, i.e. that its basic properties (mean, variance, covariances and correlations) do not depend on time.

As explained above, if the original time series is already stationary, one has $d = 0$. If the original series y_t shows a linear trend, then $z_t = y_t - y_{t-1}$ will be stationary, in other words $d = 1$. If the original series shows a quadratic trend, then $d = 2$. Other values are rare, since a behaviour which does not fit into any of the three previous categories typically indicates that the data should either be pre-treated in a more advanced way (e.g. if the original series show an exponentially increasing or decreasing trend, one should first take $z_t = log(y_t)$), or that the data is dominated by noise.

To test for stationarity, one can use the Dickey-Fuller test [Dickey and Fuller, 1979]. The Dickey-Fuller test is a so-called unit root test. It can be seen that, when the polynomial $\Phi(B)$ has d unit roots (i.e., $\Phi(B) = 0$ for $B = 1$ with multiplicity d), then the time series expressed in terms of $z_t = (1 - B)^d y_t$ will be stationary. Note that we in fact use an Augmented Dickey-Fuller test (ADF), which is an extension of the standard Dickey-Fuller test to a wider range of Time Series models.

In this ADF test, the null hypothesis is that a unit root is present in the time series. A small p-value (typically: smaller than 0.05) indicates a strong rejection of this null hypothesis, and is therefore indicative of a stationary time series.

As we will discuss below, we indeed find small p-values for all the cases that we have looked at. In other words $d = 0$, i.e.: the time series that we study are stationary without any transformation.

3.1.2 Predictions

One of the main applications of ARIMA models is to make predictions for the (near) future. It can be shown that the variance on a predicted value m time lags ahead of the most recent observation is

$$\text{Var}(m) = \sigma_\epsilon^2 \sum_{j=0}^{m-1} \Psi_j^2$$

where Ψ_j represent the weights in the infinite-order causal series development of the ARIMA model into an equivalent pure MA model. The main point here is that this sum is strictly increasing (therefore the variance obviously increases as one tries to make predictions further ahead) and in fact tends towards the total variance of the complete observed time series. 95% prediction intervals are given by

$$y_{n+m} \pm 1.96\sqrt{\text{Var}(m)}$$

where y_{n+m} is the point-like estimate obtained from the ARIMA model with the n most recent observations. We will plot and briefly comment on these prediction intervals for the most interesting cases in the Results section.

3.2 GARCH

In financial models, time-varying volatility is often studied by (generalized) autoregressive conditional heteroskedasticity or (G)ARCH. ARCH models were introduced in 1982 by Engle [Engle, 1982] to describe inflation in the UK, and generalized (GARCH) by Bollerslev in 1986 [Bollerslev, 1986]. Their first application to tourism dates from 2005 [Chan et al., 2005].

The starting point is that it is often not a good assumption to assume that the prediction errors ϵ_t are independent random variables with constant variance σ_ϵ^2.

In cases where this assumption is indeed violated, it is said that the distribution presents heteroscedasticity. One can then start by replacing the previous assumption and instead writing:

$$\epsilon_t = \sigma_t e_t$$

where σ_t now represents the variance conditional on all previous information F, i.e.

$$\sigma^2 = E\left[\epsilon_t \middle| F_{t-1}\right]$$

and $e_t \sim \mathcal{N}(0,1)$, i.e. a sequence of standardized zero-mean error terms. A GARCH*(r,s)* model for the variance σ_t^2 of a time series can be written

$$\sigma_t^2 = \omega + \sum_{i=1}^{s} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2$$

where the α_i are the ARCH coefficients, and β_j the GARCH coefficients.

Note that, although from the plots of the data it seems intuitively obvious that a volatility model is relevant, because the variance is far from constant, part of the work is precisely to confirm this intuition. Let us therefore briefly describe the different steps that we have undertaken w.r.t. the GARCH model.

3.2.1 Autocorrelation analysis of the ARIMA residuals

First, we analyse the ACF (AutoCorrelation Function) and PACF (Partial AutoCorrelation Function) of the (squares of the) residuals from the ARIMA models. These should be small in order to confirm that the ARIMA models provide a good fit for the different time series. A Portmanteau Q-test (related to the Ljung-Box text) and a Lagrange Multiplier test are carried out to identify whether these residuals indeed present heteroscedasticity. These tests analyse different aspects of heteroscedasticity and are therefore complementary: as soon as one of them is positive (i.e., has a p-value <0.05), this shows that a GARCH model is indeed indicated, even if the other is inconclusive.

3.2.2 GARCH selection criteria

The selection criteria for the GARCH models depend on the AIC (Akaike Information Criterion). The AIC is a measure for the information lost by the model w.r.t. the original data, and should therefore be as small as possible. In other words, for each region of study (Spain or Autonomous Community), the corresponding GARCH model is the one that has the lowest AIC amongst all candidate models.

Once the model has been selected, the corresponding coefficients are calculated, and each coefficient receives a p-value in the usual way: the lower the p-value, the higher the rejection of the null hypothesis (namely, that the corresponding coefficient is zero), and hence the higher the significance for a non-zero coefficient with the actual value obtained.

3.2.3 Autocorrelation analysis of the GARCH residuals

After the GARCH model has been determined, we analyse the ACF (AutoCorrelation Function) and PACF (Partial AutoCorrelation Function) of the (squares of the) residuals from the actual GARCH models. These should be small in order to confirm that the GARCH models provide good a fit for the volatility in the different time series. We test explicitly whether the heteroscedasticity of the variance has been satisfactorily incorporated in the GARCH model, through a Ljung-Box test for different values of the statistic Q. We also test whether the residuals of the GARCH model fulfill normality through both a Jarque-Bera and a Shapiro-Wilk test, two classical normality tests.

Finally, we use the GARCH model obtained to make a 12-month forecast for the variance.

3.3 Data & variable

We use data from Egatur (Encuesta de gasto turístico - Survey of touristic expenditure), the statistical operation which collects data on the expenses made by non-resident visitors in Spain.

We cover the period 2008-2018. The Egatur data until 2015 were obtained from Tourspain/Turespaña, Spain's State Secretary of Tourism. Since 2015, the Egatur data are collected by INE (Instituto Nacional de Estadística), the official Spanish Statistical Office. Note that for the data until 2015, we have access to the full daily data obtained directly from the surveys, whereas since 2015 we currently have access only to the monthly summaries published by INE.

The variable that we analyse is the average daily expenditure (DE) per (foreign) tourist. This is obtained from the Egatur data as

$$\mathrm{DE} = \frac{\text{total expenditure during visit}}{\text{number of overnight stays}}$$

where both quantities are measured per tourist. A summary of the data is given in Table 1.

A histogram of our variable, namely the daily expenditure by international tourists, general for Spain, based on these data can be found in Fig. 1. The size of the bars in this histogram are therefore proportional to the number of months in which the daily average expenditure lied within the corresponding range of values.

We have performed analyses both for Spain in general, and for the different Autonomous Communities separately. The most interesting results for the Autonomous Communities most relevant for Spain's tourism industry will be shown and discussed explicitly in the Results section below. In particular, we have selected the Autonomous Communities of our focus based on a simple indicator, namely the ones with the largest number of tourist arrivals in the period October 2015–August 2018 as indicated in the below table 2 (source: INE 2018).

	mean	sd	min	max	range
España	182.57	19.77	138.80	235.20	96.41
Andalucia	168.74	40.90	100.78	374.20	273.42
Aragón	174.04	49.93	97.43	374.17	276.74
Asturias	168.89	60.27	78.29	491.20	412.91
Baleares	175.58	65.31	68.36	474.90	406.54
Canarias	172.79	55.66	60.35	412.96	352.61
Cantabria	155.85	37.83	89.22	350.73	261.52
Cataluña	206.75	46.82	96.57	328.39	231.83
Castilla y León	167.29	38.41	100.49	288.51	188.02
Castilla la Mancha	160.13	41.50	92.47	346.62	254.15
Extremadura	144.45	34.54	65.86	237.44	171.58
Galicia	160.41	49.44	85.46	489.00	403.54
Madrid	241.07	52.58	110.64	354.97	244.33
Murcia	135.41	42.99	61.94	399.62	337.69
Navarra	185.30	64.25	63.78	445.56	381.77
Pais Vasco	209.18	54.99	120.89	491.59	370.70
Rioja	173.52	72.96	58.77	413.96	355.19
Valencia	153.75	30.20	103.34	268.75	165.41
Ceuta	203.26	161.62	57.46	530.14	472.68
Melilla	130.00	96.68	37.43	333.94	296.50

Table 1: Summary of Data: mean, standard deviation, minimum, maximum and range of the daily expenditure (in €) per international tourist in Spain and in the Spanish Autonomous Communities, 2008–2018

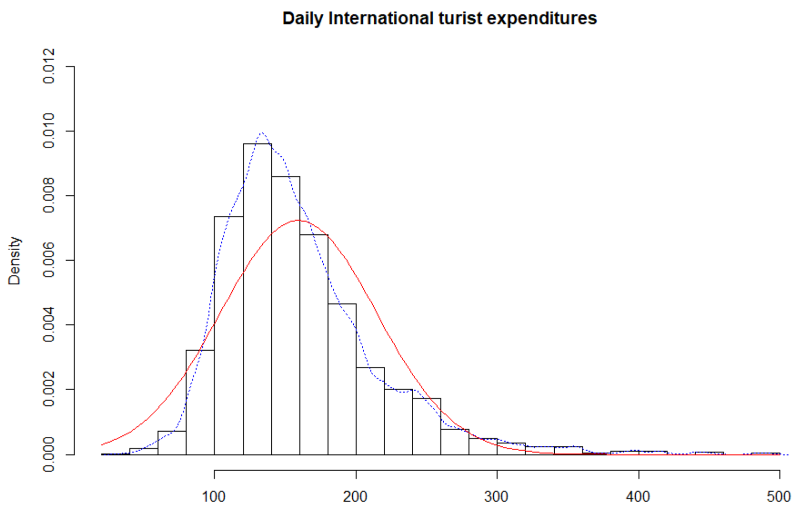

Figure 1: Histogram for the average daily expenditure by international tourists in Spain, measured per month. The blue line shows the density curve, the red line is the normal distribution approximation.

Autonomous Community	absolute number	with respect to total for Spain (%)
Catalonia	1.546.953	23,73
Canary Islands	1.135.416	17,42
Balearic Islands	1.090.633	16,73
Andalusia	917.669	14,08
Valencian Community	695.375	10,67
Madrid Community	523.851	8,04
Other Autonomous Communities	609.523	9,35
Total	6.519.419	

Table 2: Number of tourist arrivals (average per month for period Oct 2015–Aug 2018)

The evolution of the average daily expenditure per tourist for the main Autonomous Communities is plotted in Fig. 2. From that plot, it should come as no surprise that we will find that most Autonomous Communities do not show any clear (upward or downward) trend but are roughly stationary. Madrid (blue) and Catalonia (yellow) are the exceptions, they are the only two regions with a clear upward trend in the past few years. This becomes even clearer when plotting the data separately for Spain and for the 6 Autonomous Communities that we have identified earlier, see Fig. 3 and 4.

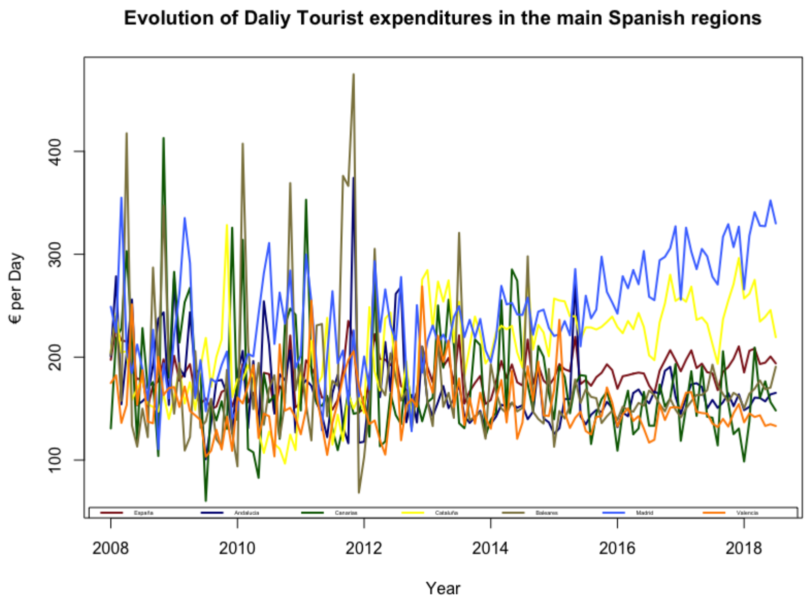

Figure 2: Time series plot showing the evolution of the daily tourist expenditure in the Spanish Autonomous Communities most important for tourism. The blue line corresponds to Madrid, the yellow line to Catalonia. Other Autonomous Communities of interest are discussed in the main text.

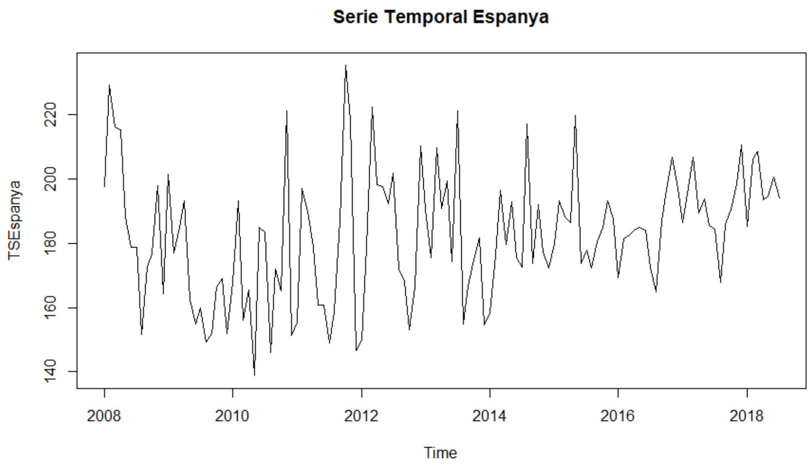

Figure 3: Time series for Spain

4 Results & analysis

4.1 Unit root test

As mentioned above, we perform an Augmented Dickey-Fuller (ADF) or unit root test. The absence of unit roots in all cases (both the global results for Spain and the individual Autonomous Communities separately) means that the original time series are already approximately stationary, without the need for any transformation. The results are summarized in Table 3. Note that the ADF statistic is a negative number, which is compared to the critical value (which depends on the sample size and a number of other parameters) to obtain the p-value. As usual, the null hypothesis H_0 (namely that there is a unit root) is rejected for $p < 0.05$, in which case the time series can be considered stationary.

	ADF statistic	p-value
Spain	-54158	<0.01
Catalonia	-43424	<0.01
Canary Islands	-50554	<0.01
Andalusia	-37042	0.02671
Valencian Community	-35114	0.04413
Madrid Community	-41118	<0.01

Table 3: ADF test results

We plot the case of Spain for illustrative purposes, see Fig. 5. Note that, although several roots lie close to the complex unit circle, detailed analysis shows that they all lie strictly within the circle, thereby confirming the approximately stationary character of the time series.

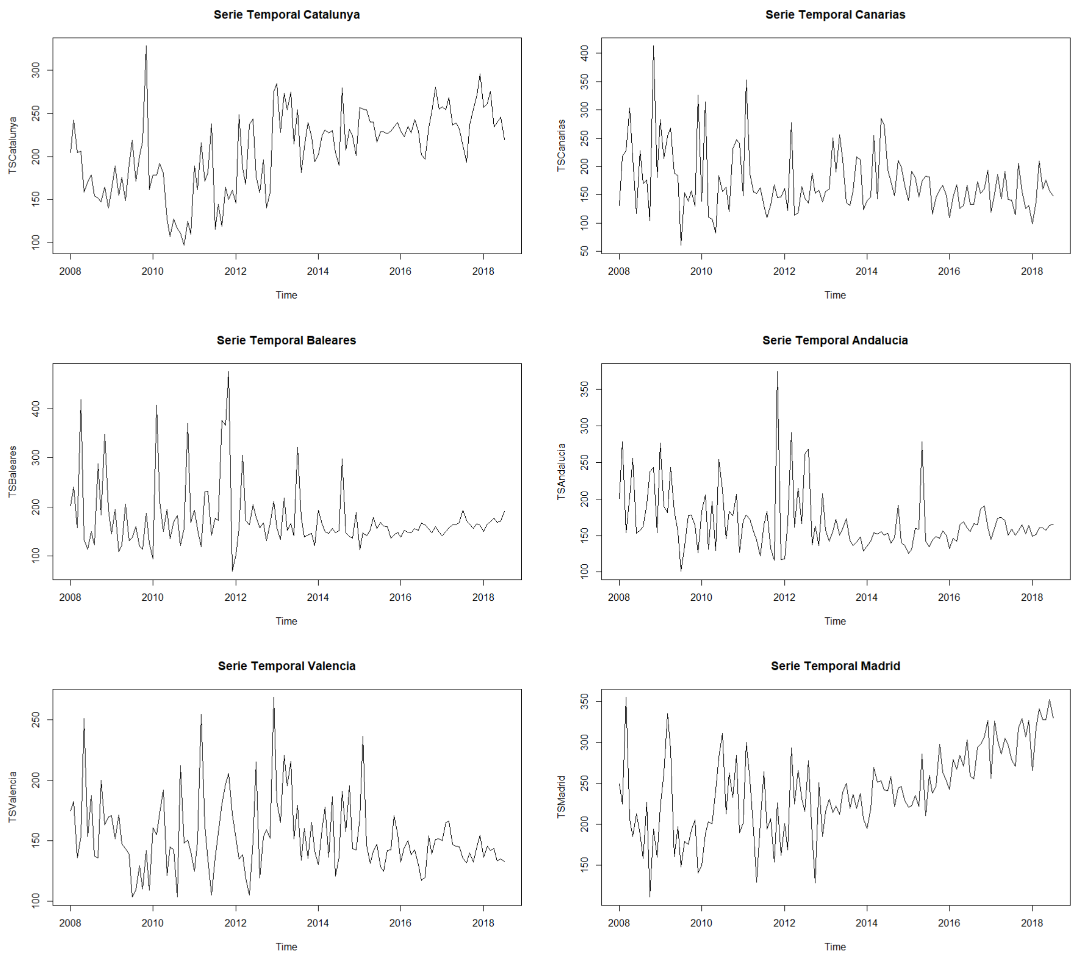

Figure 4: Time series for the Autonomous Communities with highest number of tourists. From left to right, row by row: Catalonia and Canary Islands; Balearic Islands and Andalusia; Valencia and Madrid. Note the upward trend for Catalonia and Madrid.

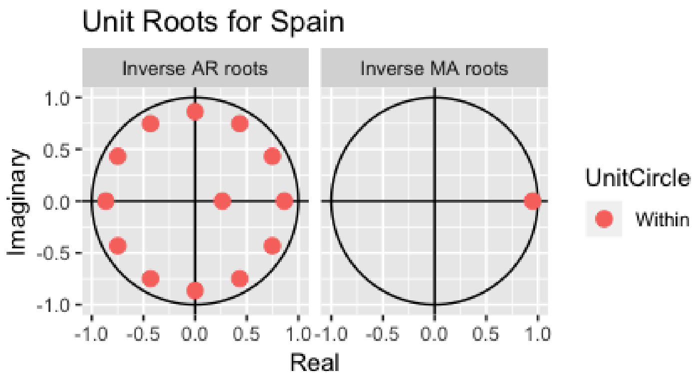

Figure 5: Unit root test, global data (Spain). All complex roots lie within the unit circle. The time series is therefore stationary.

4.2 ARIMA models

We use R to perform an ARIMA modelling for the global data for Spain, as well as for each Autonomous Community separately. The results are summarized in Table 4. Note that these models are selected based on the AIC criterion (which should be minimum) and a maximum likelihood criterion.

	ARIMA model
Spain	(1,1,1)(1,0,0)[12]
Catalonia	(2,1,1)
Canary Islands	(2,1,1)(1,0,0)[12]
Balearic Islands	(0,1,1)
Andalusia	(0,1,1)
Valencian Community	(1,0,0)
Madrid Community	(2,1,1)(1,0,1)[12]

Table 4: ARIMA results for Spain and the main Autonomous Communities

It is remarkable that, apart from Madrid and the Canary Islands, all the main Autonomous Communities separately are best described by non-seasonal ARIMAs. It should be stressed that we are analysing the average daily tourist expenditure, so this non-seasonality is unrelated to the question whether the total amount of tourists (and therefore the total tourist expenditure) is seasonal or not. In any case, the global data for Spain does indeed show a clear seasonal component. The estimated coefficients of all these ARIMA models, together with their standard error, normalized z-value, p-value and significance level, are given in Table 5

4.2.1 ARIMA residuals

To be confident about the correctness of the ARIMA models just obtained, we plot the residuals (i.e. the random part or error which is not included in the model), see Fig. 6. We see that both for Spain globally, and for the Autonomous Communities separately,

Region	Coefficient	Estimate	Std. Error	z value	Pr(>|z|)	significance level
Spain	ar1	0.263871	0.107256	2.4602	0.01389	*
	ma1	-0.950312	0.059718	-15.9134	< 2e-16	***
	sar1	0.173779	0.091836	1.8923	0.05845	.
Andalusia	ma1	-0.929260	0.039022	-23.814	< 2.2e-16	***
Canary Islands	ar1	-0.011311	0.092941	-0.1217	0.903134	
	ar2	0.244285	0.092801	2.6324	0.008479	**
	ma1	-0.972640	0.027249	-35.6949	< 2.2e-16	***
	sar1	0.276350	0.097746	2.8272	0.004695	**
Catalonia	ar1	0.343178	0.101467	3.3822	0.0007192	***
	ar2	0.165015	0.097990	1.6840	0.0921806	.
	ma1	-0.920066	0.047251	-19.4719	< 2.2e-16	***
Madrid	ar1	0.223616	0.096253	2.3232	0.02017	*
	ar2	0.178963	0.094747	1.8889	0.05891	.
	ma1	-0.929485	0.036092	-25.7530	< 2.2e-16	***
	sar1	0.831374	0.112619	7.3822	1.557e-13	***
	sma1	-0.610767	0.156602	-3.9001	9.615e-05	***
Valencia	ar1	0.212744	0.086624	2.4559	0.01405	*
	intercept	153.754543	3.305807	46.5104	< 2e-16	***
Balearic Islands	ma1	-0.955315	0.032544	-29.355	< 2.2e-16	***

The significance codes correspond to 0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Table 5: ARIMA coefficients for the models found in Table 4.

the residuals are indeed small, certainly well below the standard level which is typically considered in the literature. In other words the fit of the different models to the data of the different Autonomous Communities is very good. However, we also see from this figure that the residuals are not distributed evenly throughout time, which is a first, visual indication that a GARCH model for the variability might be required.

4.2.2 Predictions

80% and 95% predictions with a horizon of 12 months are given in Fig. 7 for Spain, Catalonia and Madrid. The other Autonomous Communities give predictions of little interest, because the obtained ARIMA models are too simple to make reliable predictions.

4.3 GARCH

4.3.1 ARIMA residuals and GARCH tests

The ARIMA residuals and the results from the Portmanteau Q and Lagrange Multiplier tests are shown in Figs. 9 and 9. Note that the Portmanteau Q test and the LM test separately would not be sufficient in all cases, but the combination of both is conclusive, and shows that a GARCH model to address the heteroscedasticity of the variance is indicated in all cases.

4.3.2 Selection of GARCH models

The obtained GARCH models are shown in Table 6. As in the case of the ARIMA models, the GARCH models are selected based on minimum AIC amongst all candidate models. The coefficients for these GARCH models are given in table 7

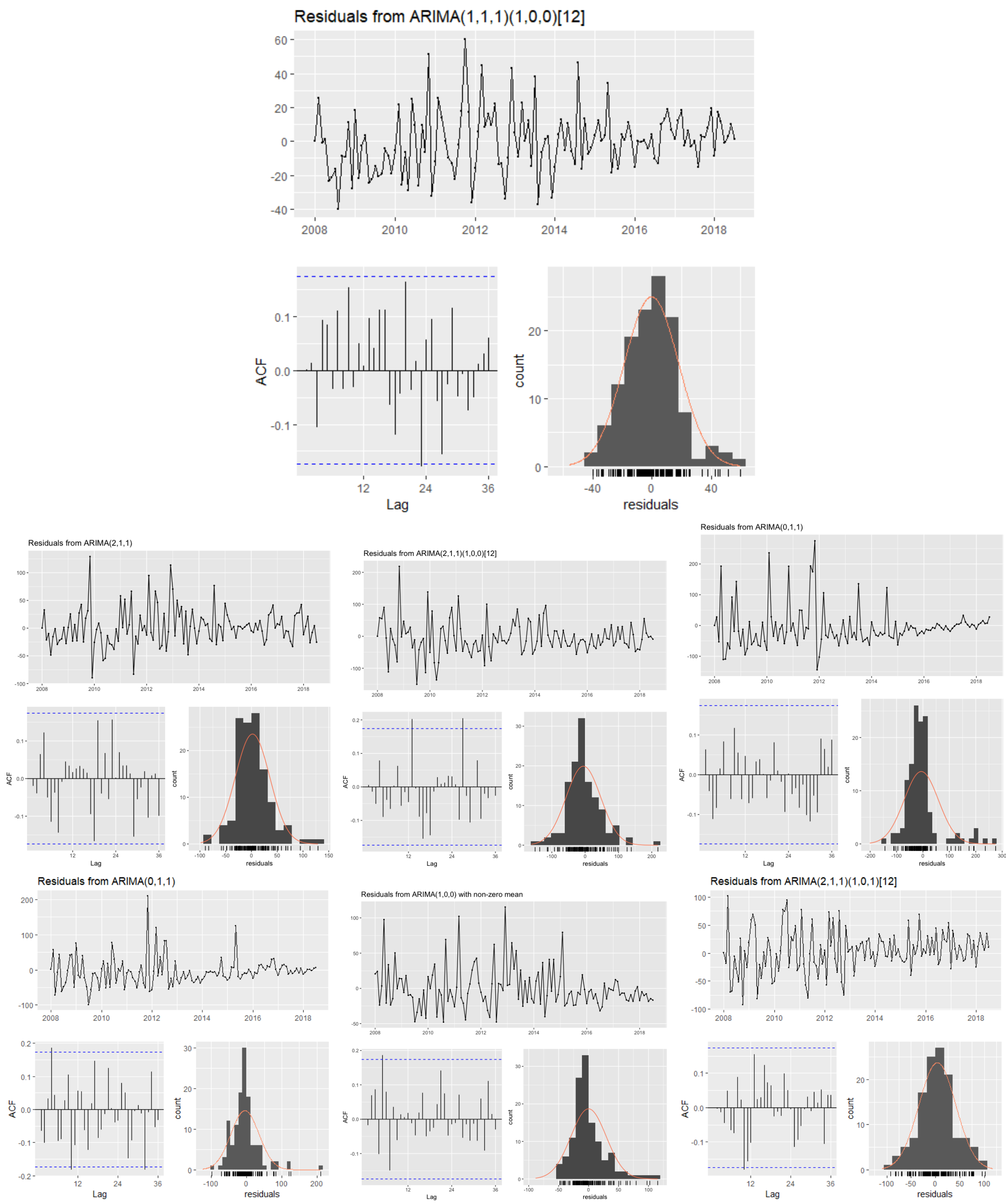

Figure 6: Residuals, ACF of the residuals, and histogram with model prediction for Spain (top) and for the main Autonomous Communities. From left to right, row by row: Catalonia, Canary Islands and Balearic Islands; Andalusia, Valencia and Madrid.

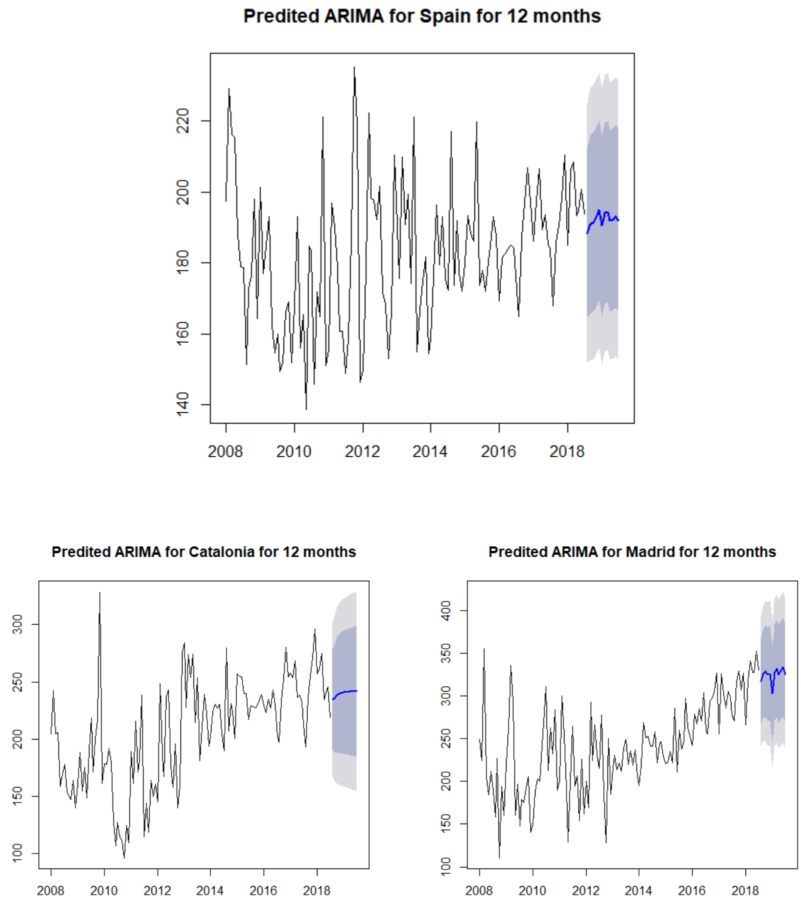

Figure 7: 12-month ARIMA predictions for Spain (top) and for Catalonia (bottom left) and Madrid (bottom right).

	GARCH model
Spain	(1,1)
Catalonia	(1,3)
Canary Islands	(1,1)
Balearic Islands	(1,5)
Andalusia	(1,3)
Valencian Community	(3,1)
Madrid Community	(1,1)

Table 6: GARCH models obtained for Spain and the Autonomous Communities of interest

Region	Coefficient	Estimate	Std. Error	t value	Pr(>|t|)	significance level
Spain	mu	1.6588831	1.4936002	1.111	0.2667	
	omega	0.0003399	7.0424689	0.000	1.0000	
	alpha1	0.0708527	0.0327158	2.166	0.0303	*
	beta1	0.9219997	0.0397499	23.195	<2e-16	***
Andalusia	mu	-1.726e+00	1.389e+00	-1.243	0.214	
	omega	6.329e+00	4.306e+01	0.147	0.883	
	alpha1	1.000e+00	2.355e-01	4.247	2.17e-05	***
	beta1	2.186e-02	3.122e-02	0.700	0.484	
	beta2	1.000e-08	(*)	(*)	(*)	
	beta3	3.763e-01	6.846e-02	5.496	3.88e-08	***
Canarias	mu	-5.15325	3.51480	-1.466	0.143	
	omega	9.11751	28.52242	0.320	0.749	
	alpha1	0.08011	0.05770	1.388	0.165	
	beta1	0.90844	0.05254	17.291	<2e-16	***
Catalonia	mu	5.997e-01	2.519e+00	0.238	0.8118	
	omega	4.178e+01	5.867e+01	0.712	0.4764	
	alpha1	3.258e-01	1.301e-01	2.503	0.0123	*
	beta1	3.619e-02	1.599e-01	0.226	0.8209	
	beta2	1.000e-08	1.026e-01	0.000	1.000	
	beta3	6.301e-01	9.662e-02	6.522	6.95e-11	***
Madrid	mu	8.20318	2.60212	3.152	0.00162	**
	omega	26.79504	33.76451	0.794	0.42744	
	alpha1	0.14624	0.08451	1.730	0.08355	.
	beta1	0.82641	0.08351	9.896	< 2e-16	***
Valencia	mu	-3.912e-01	8.689e+00	-0.045	0.964	
	omega	8.707e-04	8.277e+00	0.000	1.000	
	alpha1	1.000e-08	7.253e-01	0.000	1.000	
	alpha2	1.000e-08	5.313e-01	0.000	1.000	
	alpha3	1.016e-01	1.501e-01	0.677	0.498	
	beta1	8.907e-01	1.062e-01	8.386	<2e-16	***
Balearic Islands	mu	-6.681e+00	1.511e+00	-4.421	9.84e-06	***
	omega	4.223e-03	2.763e+01	0.000	1	
	alpha1	8.562e-01	2.063e-01	4.149	3.34e-05	***
	beta1	1.000e-08	(*)	(*)	(*)	
	beta2	1.000e-08	1.288e-02	0.000	1	
	beta3	1.000e-08	(*)	(*)	(*)	
	beta4	3.376e-01	(*)	(*)	(*)	
	beta5	7.740e-02	(*)	(*)	(*)	

The significance codes correspond to 0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Table 7: GARCH coefficients for the models found in Table 6[a)]

[a)] In the cases marked with (*), the standard error could not be evaluated due to a singularity in the covariance matrix.

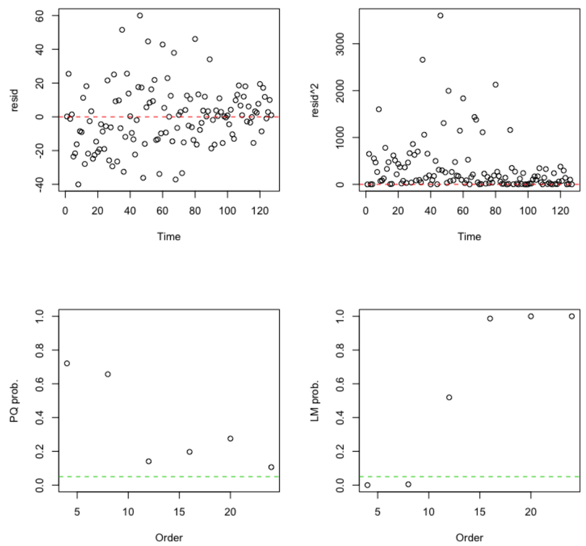

Figure 8: Analysis of the ARIMA residuals (top) and GARCH tests (bottom): Portmanteau Q test (left) and Lagrange Multiplier test (right) for Spain.

4.3.3 GARCH residuals & future predictions

As shows in Table 8, the Ljung-Box tests (both for the residuals and the residuals squared, and at different values of the test statistic Q) confirm that the GARCH have been effective in the sense that, in all cases, the p-value $p \gg 0.05$, therefore the null hypothesis that the GARCH residuals no longer present heteroscedasticity is accepted. The Jarque-Bera and the Shapiro-Wilk tests contrast the null hypothesis H_0 that these residuals are normally distributed. H_0 is accepted for Madrid, and (barely) for Spain, but rejected ($p < 0.05$) in all other cases. This indicates that in all those cases other than Spain and Madrid, a more complicated (non-linear) GARCH model will be needed. We leave this for future work.

Finally, we plot the GARCH prediction for Spain, as well as for Andalusia and the Balearic Islands, in Fig. 10.

5 Conclusions

We have developed ARIMA and GARCH models for the daily expenditure from international tourists in Spain over the period 2008–2018 and its volatility. We have studied Spain in general, as well as the Autonomous Communities most relevant for Spanish tourism. The ARIMA models are quite satisfying in general, in the sense that a non-trivial ARIMA was found for each case, which fits the data very well. Reasonable predictions are obtained in some cases, less so in others.

The results for the GARCH models should be interpreted as work in progress: although we find GARCH models that solve the heteroscedasticity of the ARIMA residuals satisfactorily, these residuals are still strongly non-normal in many cases, indicated that

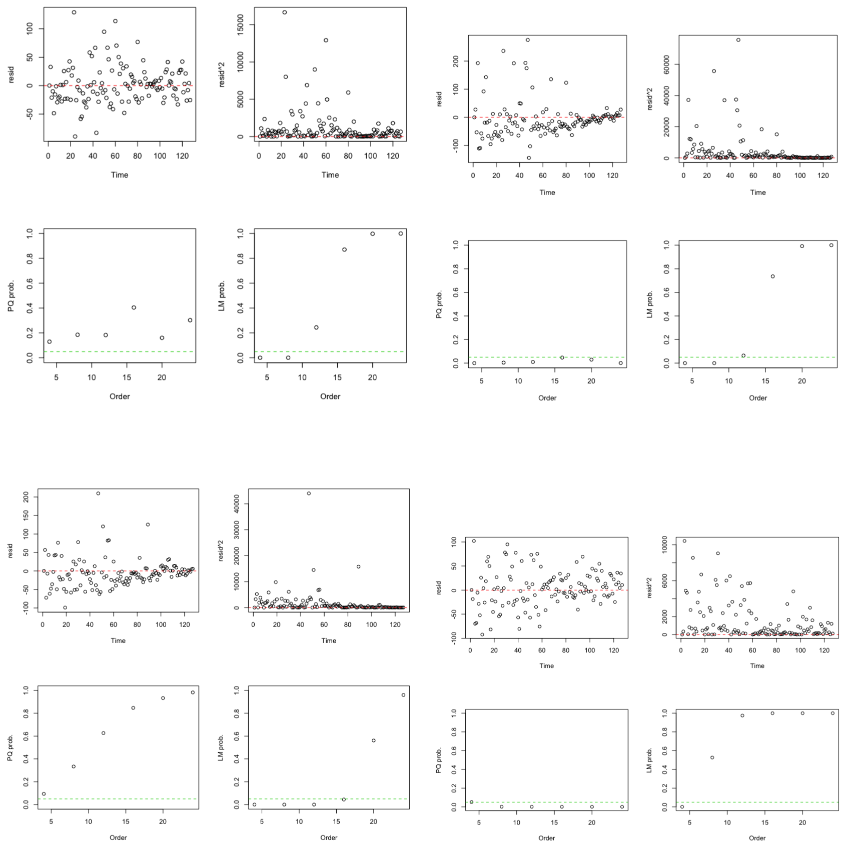

Figure 9: Analysis of the ARIMA residuals (top) and GARCH tests (bottom): Portmanteau Q test (left) and Lagrange Multiplier test (right) for (row by row) Catalonia and Balearic Islands; Andalusia and Madrid.

Region	Test	R/R^2	Test Statistic	value	p-value
Spain	Jarque-Bera Test	R	Chi^2	5.756262	0.05623979
	Shapiro-Wilk Test	R	W	0.9801071	0.05855171
	Ljung-Box Test	R	Q(10)	8.414394	0.588428
		R	Q(15)	13.88077	0.53459
		R	Q(20)	19.81559	0.4695176
		R^2	Q(10)	6.260922	0.792886
		R^2	Q(15)	11.44333	0.7205816
		R^2	Q(20)	14.22665	0.8188208
Catalonia	Jarque-Bera Test	R	Chi^2	8.310938	0.01567844
	Shapiro-Wilk Test	R	W	0.9506902	0.0001548937
	Ljung-Box Test	R	Q(10)	8.463025	0.5837051
		R	Q(15)	9.525819	0.8484608
		R	Q(20)	15.14329	0.7681507
		R^2	Q(10)	5.39433	0.8633295
		R^2	Q(15)	7.2839	0.9492875
		R^2	Q(20)	13.2644	0.8657536
Canary Islands	Jarque-Bera Test	R	Chi^2	19.75103	5.141847e-05
	Shapiro-Wilk Test	R	W	0.9615736	0.001157347
	Ljung-Box Test	R	Q(10)	3.131791	0.9781501
		R	Q(15)	7.953232	0.925644
		R	Q(20)	12.30734	0.9050914
		R^2	Q(10)	7.145179	0.7116718
		R^2	Q(15)	14.51513	0.4868731
		R^2	Q(20)	15.21589	0.7639239
Balearic Islands	Jarque-Bera Test	R	Chi^2	57.21383	3.768097e-13
	Shapiro-Wilk Test	R	W	0.8959867	6.309865e-08
	Ljung-Box Test	R	Q(10)	8.355261	0.5941784
		R	Q(15)	11.19847	0.7384024
		R	Q(20)	12.78524	0.886407
		R^2	Q(10)	3.402698	0.9702994
		R^2	Q(15)	4.829972	0.9934707
		R^2	Q(20)	7.686855	0.9937451
Andalusia	Jarque-Bera Test	R	Chi^2	289.0935	0
	Shapiro-Wilk Test	R	W	0.8778514	8.237248e-09
	Ljung-Box Test	R	Q(10)	6.899207	0.734925
		R	Q(15)	11.99736	0.6792286
		R	Q(20)	16.26067	0.700327
		R^2	Q(10)	2.546925	0.9901741
		R^2	Q(15)	3.939372	0.9979291
		R^2	Q(20)	8.454818	0.9884051
Madrid	Jarque-Bera Test	R	Chi^2	3.413354	0.1814678
	Shapiro-Wilk Test	R	W	0.9845099	0.1574786
	Ljung-Box Test	R	Q(10)	6.833238	0.741089
		R	Q(15)	11.98993	0.6797914
		R	Q(20)	15.51412	0.7462733
		R^2	Q(10)	14.07458	0.1696171
		R^2	Q(15)	20.20095	0.1643587
		R^2	Q(20)	23.77055	0.2525586
Valencia	Jarque-Bera Test	R	Chi^2	61.11834	5.351275e-14
	Shapiro-Wilk Test	R	W	0.9128318	5.112739e-07
	Ljung-Box Test	R	Q(10)	13.90641	0.1773034
		R	Q(15)	15.16975	0.4392602
		R	Q(20)	17.499	0.6203744
		R^2	Q(10)	3.097789	0.9790281
		R^2	Q(15)	5.89876	0.9813986
		R^2	Q(20)	6.995376	0.9967002

Table 8: GARCH residuals tests

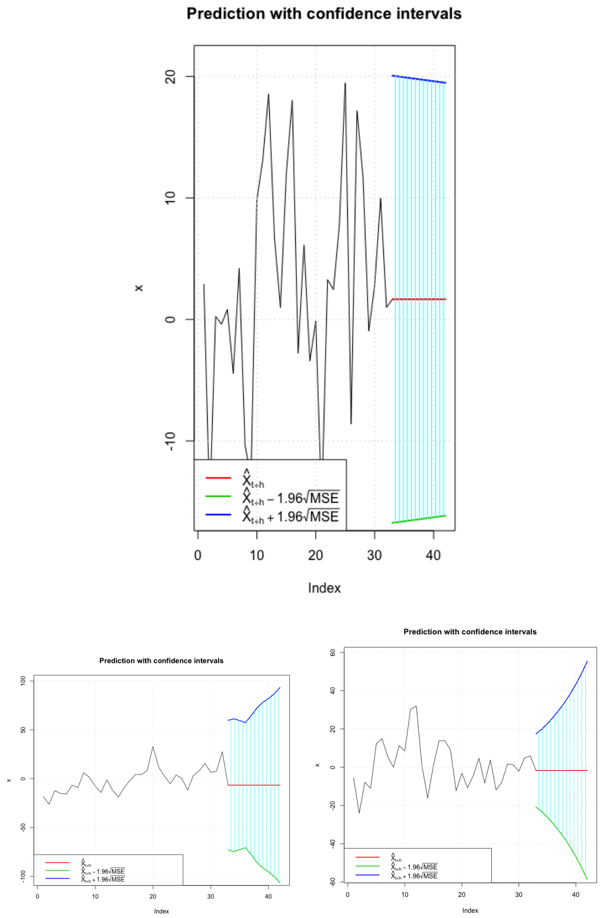

Figure 10: GARCH predictions for Spain (top), the Balearic Islands (bottom left) and Andalusia (bottom right).

more complicated (non-linear) GARCH models might be needed, such as EGARCH or NGARCH models.

However, we believe that part of the complications that we have run into might be due to the dichotomy in the data: the data before 2015 were obtained from the original EGATUR surveys, whereas the data starting from 2015 is taken from the monthly summaries published by INE. This might explain, at least partially, why we observe a lower volatility in the period since 2015 than in the period closer to 2008. On the other hand, the lower volatility seems to set in already in 2013, so perhaps the lower volatility is a true effect of the mitigation of the 2008 crisis.

We hope to be able to answer this question, and to generally improve on the results presented here, by obtaining and analysing the full EGATUR surveys for these more recent data (since 2015) too.

References

UNWTO Tourism Highlights, 2018 (Madrid).

WTTC - World Travel and Tourism Council: Travel & Tourism Economic Impact, 2018.

Aguiló, E., Alegre, J., and Sard, M. (2005). The persistence of the sun and sand tourism model. *Tourism Management*, 26(2):219 – 231.

Athanasopoulos, G., Hyndman, R. J., Song, H., and Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, 27(3):822 – 844. Special Section 1: Forecasting with Artificial Neural Networks and Computational Intelligence Special Section 2: Tourism Forecasting.

Bartolomé, A., McAleer, M., Ramos, V., and Rey-Maquieira, J. (2009). Modelling air passenger arrivals in the balearic and canary islands, spain. *Tourism Economics*, 15(3):481–500.

Bernier, E. T., Sánchez, R. R., and Díaz, B. R. (2014). La crisis económica en el sector turístico. Un análisis de sus efectos en la Costa del Sol. *Revista del Analisis Turistico*, 18.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307 – 327.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2013). *Time Series Analysis (4th ed.)*. Wiley.

Chan, F., Lim, C., and McAleer, M. (2005). Modelling multivariate international tourism demand and volatility. *Tourism Management*, 26(3):459 – 471.

Chang, C.-L., Khamkaew, T., Tansuchat, R., and McAleer, M. (2011). Interdependence of international tourism demand and volatility in leading asean destinations. *Tourism Economics*, 17(3):481–507.

Chang, C.-L., McAleer, M., and Slottje, D. (2009). Chapter 11 modelling international tourist arrivals and volatility: An application to taiwan. *Contributions to Economic Analysis*, 288.

Chen, C.-F., Lai, M.-C., and Yeh, C.-C. (2012). Forecasting tourism demand based on empirical mode decomposition and neural network. *Knowledge-Based Systems*, 26:281 – 287.

Claveria, O. and Torra, S. (2014). Forecasting tourism demand to catalonia: Neural networks vs. time series models. *Economic Modelling*, 36:220 – 228.

Claveria, O., Torra, S., and Monte, E. (2016). Modelling tourism demand to Spain with machine learning techniques. The impact of forecast horizon on model selection. *Revista de Economía Aplicada*, XXIV(72).

Coshall, J. T. (2009). Combining volatility and smoothing forecasts of uk demand for international tourism. *Tourism Management*, 30(4):495 – 511.

Cunado, J., Gil-Alana, L., and de Gracia, F. P. (2005). The nature of seasonality in spanish tourism time series. *Tourism Economics*, 11(4):483–499.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a):427–431.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

Fernandes, P., Teixeira, J., Ferreira, J., and Azevedo, S. (2008). Modelling tourism demand: a comparative study between artificial neural networks and the Box-Jenkins methodology. *Journal for Economic Forecasting*, 5(3):30–50.

García-Ferrer, A. and Queralt, R. A. (1997). A note on forecasting international tourism demand in spain. *International Journal of Forecasting*, 13(4):539 – 549.

Garin-Munoz, T. and Amaral, T. P. (2000). An econometric model for international tourism flows to spain. *Applied Economics Letters*, 7(8):525–529.

Goh, C. and Law, R. (2011). The methodological progress of tourism demand forecasting: A review of related literature. *Journal of Travel & Tourism Marketing*, 28(3):296–317.

González, P. and Moral, P. (1995). An analysis of the international tourism demand in spain. *International Journal of Forecasting*, 11(2):233 – 251.

Hoti, S., León, C. J., and McAleer, M. International Tourism Demand and Volatility Models for the Canary Islands.

Liu, J. and Sriboonchitta, S. (2013). Analysis of volatility and dependence between the tourist arrivals from china to thailand and singapore: A copula-based garch approach. In Huynh, V.-N., Kreinovich, V., Sriboonchitta, S., and Suriya, K., editors, *Uncertainty Analysis in Econometrics with Applications*, pages 283–294, Berlin, Heidelberg. Springer Berlin Heidelberg.

Makoni, T. and Chikobvu, D. (2018). Modelling tourism demand volatility using a seasonal autoregressive integrated moving average autoregressive conditional heteroscedasticity model for victoria falls rainforest arrivals in zimbabwe. *Journal of Economic and Financial Sciences*, 11(1):1–9.

Martins, L. F., Gan, Y., and Ferreira-Lopes, A. (2017). An empirical analysis of the influence of macroeconomic determinants on world tourism demand. *Tourism Management*, 61:248 – 260.

Nicolau, J. L. and Más, F. J. (2005). Heckit modelling of tourist expenditure: evidence from spain. *International Journal of Service Industry Management*, 16(3):271–293.

Palmer, A., Montaño, J. J., and Sesé, A. (2006). Designing an artificial neural network for forecasting tourism time series. *Tourism Management*, 27(5):781 – 790.

Papatheodorou, A. and Song, H. (2005). International tourism forecasts: Time-series analysis of world and regional data. *Tourism Economics*, 11(1):11–23.

Perez, E. A. and Juaneda, S. C. (2000). Tourist expenditure for mass tourism markets. *Annals of Tourism Research*, 27(3):624 – 637.

Petrevska, B. (2017). Predicting tourism demand by a.r.i.m.a. models. *Economic Research-Ekonomska Istraživanja*, 30(1):939–950.

Rodriguez, X. A. and Rivadulla, R. (2012). Tourism in Spain: Disaggregated Analysis of the International Demand. *Regional and Sectoral Economic Studies*, 12(1):91–98.

Shareef, R. and McAleer, M. (2005). Modelling international tourism demand and volatility in small island tourism economies. *International Journal of Tourism Research*, 7(6):313–333.

Sheldon, P. J. (1993). Forecasting tourism: Expenditures versus arrivals. *Journal of Travel Research*, 32(1):13–20.

Song, H. and Li, G. (2008). Tourism demand modelling and forecasting—a review of recent research. *Tourism Management*, 29(2):203 – 220.

Song, H., Li, G., Witt, S. F., and Athanasopoulos, G. (2011). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27(3):855 – 869. Special Section 1: Forecasting with Artificial Neural Networks and Computational Intelligence Special Section 2: Tourism Forecasting.

Song, H., Li, G., Witt, S. F., and Fei, B. (2010). Tourism demand modelling and forecasting: How should demand be measured? *Tourism Economics*, 16(1):63–81.

Tsay, R. S. (2000). Time series and forecasting: Brief history and future research. *Journal of the American Statistical Association*, 95(450):638–643.

Tularam, G. A., Wong, V. S. H., and Nejad, S. A. S. (2012). Modeling Tourist Arrivals Using Time Series Analysis: Evidence From Australia. *Journal of Mathematics and Statistics*, 8:348 – 360.

WB (2018). World Tourism Organization, Yearbook of Tourism Statistics, Compendium of Tourism Statistics and data files. Data accessed on the World Bank's website, https://data.worldbank.org, 24/09/2018.