Journal of Econometrics 147 (2008) 372–383

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

An alternative approach to estimating demand: Neural network regression with conditional volatility for high frequency air passenger arrivals

Marcelo C. Medeiros [a,*], Michael McAleer [b], Daniel Slottje [c], Vicente Ramos [d], Javier Rey-Maquieira [d]

[a] Department of Economics, Pontifical Catholic University of Rio de Janeiro, Brazil
[b] School of Economics and Commerce, University of Western Australia, Australia
[c] Department of Economics, Southern Methodist University, United States
[d] Department of Applied Economics, University of the Balearic Islands, Spain

ARTICLE INFO

Article history:
Available online 18 September 2008

JEL classification:
D12
C14
C22

Keywords:
Semi-parametric models
Neural networks
Smooth transition
Nonlinear models
Time series
Passenger arrivals
Tourism demand

ABSTRACT

In this paper we provide an alternative approach to analyze the demand for international tourism in the Balearic Islands, Spain, by using a neural network model that incorporates time-varying conditional volatility. We consider daily air passenger arrivals to Palma de Mallorca, Ibiza and Mahon, which are located in the islands of Mallorca, Ibiza and Menorca, respectively, as a proxy for international tourism demand for the Balearic Islands. Spain is a world leader in terms of total international tourist arrivals and receipts, and Mallorca is one of the most popular destinations in Spain. For tourism management and marketing, it is essential to forecast high frequency international tourist demand accurately. As it is important to provide sensible international tourism demand forecast intervals, it is also necessary to model their variances accurately. Moreover, time-varying variances provide useful information regarding the risks associated with variations in international tourist arrivals.

1. Introduction

Most of the articles in the literature of demand systems deal with a utility-based (or parametric) approach to modelling consumer demand. The traditional approach to quantifying consumer preferences assumes the existence of either an explicit or implicit direct (or indirect) utility function, uncovers the system of demand functions through the optimization process (or simply specifies them), and then tests a specified functional form of demand functions against relevant data to determine if the Slutsky properties implied by the theory (such as non-negativity, symmetry, homogeneity of degree zero in prices and income) are found to hold (see, for example, Basmann et al. (1987)).

Such an approach is valuable and necessary if the intent of the inquiry is to determine a specific characteristic (or characteristics) of the economic agent or variable under scrutiny. That is, estimation of the own-price elasticity of demand of a commodity may be the primary reason for the analysis. Concurrently, the validity of the specification of the functional form itself may be the chief point of inquiry. Economists undertake these exercises because they continue to search for a better understanding of how consumer preferences are formulated and how consumer demand is impacted by interventions such as changes in income, relative prices and demographics. The functional form is also useful in the construction of true cost of living indices, and thereby in allowing for predictions about how the true cost of living may change when prices and income change (see, for example, Basmann and Slottje (1987)). Most of the papers in the literature present different (and innovative methods) to explore such challenging questions.

There are, of course, other objectives of consumer demand studies and other approaches to analyzing consumer demand. The primary objective of demand analysis may be to forecast consumer demand for a given activity or commodity over time, and to determine what factor(s), if any, may impact the demand for that activity or commodity in the future. While it is possible to take traditional systems of demand functions and to estimate the parameters of the system-wide model (see Theil (1975, especially pp. 322–326)), and then to use those coefficient estimates for prediction and policy purposes, there are other ways in which to accomplish these aims. In some situations, the nature of the data themselves may present problems with using a standard approach to estimating consumer demand. If the data are high frequency

* Corresponding author.
E-mail addresses: mcm@econ.puc-rio.br (M.C. Medeiros), michael.mcaleer@gmail.com (M. McAleer), dslottje@smu.edu (D. Slottje), vicente.ramos@uib.es (V. Ramos), javier.rey@uib.es (J. Rey-Maquieira).

0304-4076/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2008.09.018

with non-linearities, traditional systems of demand equations may be ill-equipped to deal with the potentially complicated estimation issues.

This paper shows that the use of statistical (non-parametric) models allows us to exploit the potentially rich vein of information that may be inherent in particular high frequency data under scrutiny by specifying explicitly the volatility inherent in the data over time, and to discover potential changes in consumer behavior based on the estimates of volatility that are extracted from the statistical model. The agnostic view (with respect to the functional form of the demand equations) taken by a statistical model of consumer demand may have other advantages over the traditional approach. Such models not only avoid the integrability problem (see Samuelson (1947)), but can be interpreted as what might be called "super-reduced form" equations and, therefore, avoid some, if not all, of the usual turmoil and debate about identifiability and specification. In this paper, we show that consumer demand for air travel can be forecasted effectively using a neural network model with conditional heteroskedastiticity that will be explained below.

Recent years have witnessed a vast development of nonlinear time series techniques (see, for example, Tong (1990) or Granger and Teräsvirta (1993)). From a parametric point of view, the Smooth Transition (Auto-)Regression, ST(A)R, proposed by Chan and Tong (1986), who called the model Smooth Threshold Auto-regression, and further developed by Luukkonen et al. (1988) and Teräsvirta (1994), has found a number of successful applications (see van Dijk et al. (2002) for a review). In the time series literature, the STAR model is a natural generalization of the Threshold Autoregressive (TAR) models, as pioneered by Tong (1978) and Tong and Lim (1980).

On the other hand, nonparametric models that do not make assumptions about the parametric form of the functional relationship between the variables to be modelled have become widely applicable due to computational advances (Härdle, 1990; Härdle et al., 1997; Fan and Yao, 2003). Another class of models, namely flexible functional forms, offers an alternative that leaves the functional form of the relationship partially unspecified. While these models do contain parameters, sometimes a very large number, the parameters are not globally identified. Identification, if achieved, is local at best, without imposing restrictions on the parameters. Usually, the parameters are not directly interpretable, as in parametric models. In most cases, these models are interpreted as nonparametric sieve (or series) approximations (Chen and Shen, 1998).

The neural network (NN) model is a prominent example of such a flexible functional form. Although the NN model can be interpreted as a parametric alternative (Kuan and White, 1994; Trapletti et al., 2000; Medeiros et al., 2006), its use in applied work is generally motivated by the mathematical result stating that, under mild regularity conditions, an NN model is capable of approximating any Borel-measurable function to any given degree of accuracy (see, for instance, Hornik et al. (1990), Gallant and White (1992), and Chen and White (1998)). NN models can also be interpreted as smooth transition models with flexible transition variables, as in Medeiros and Veiga (2005) or Medeiros et al. (2006).

This paper considers the case of NN models with GARCH (Generalized Autoregressive Conditional Heteroskedastic) errors (Engle, 1982; Bollerslev, 1986) (see also Poon and Granger (2003), McAleer (2005), and Andersen et al. (2006) for useful reviews of conditional variance models). Conditions for weak consistency and asymptotic normality of the quasi-maximum-likelihood estimator are derived. We apply the NN–GARCH model to describe the dynamics of daily air passenger arrivals at three different airports in the Balearic Islands, Spain. Our theoretical results are new and extend the work of Francq and Zakoïan (2004).

The number of air passenger arrivals is closely related to the number of tourists, as documented in Bartolomé et al. (2008). International tourism demand is important for many countries worldwide because of the tourist export receipts that they generate. Spain is one of the most visited countries in the world by international tourists, being second to France in terms of total of international tourist arrivals, and second to the USA in terms of international tourism receipts. Of the five major tourist regions in Spain, Mallorca in the Balearic Islands is one of the most popular destinations. It is clear that international tourist arrivals are important globally, as well as nationally for Spain. For purposes of tourism management and marketing, it is essential to be able to forecast the demand for tourism, as proxied by tourist arrivals, and their percentage changes accurately. As it is important to provide sensible tourist forecast intervals in addition to the forecasts themselves, it is also necessary to model the variances of the forecasts accurately.

Forecasting international tourism and their associated volatility has been considered previously in Chan et al. (2005) and Hoti et al. (2007) at the multivariate level, and in Shareef and McAleer (2007) at the univariate level. Bartolomé et al. (2008) considered different volatility models for passenger arrivals in five different airports in the Balearic Islands, including Mallorca. These papers have shown the importance and usefulness of both univariate and multivariate conditional volatility models, when used in conjunction with time series models of international tourist arrivals and their respective rates of growth.

The remainder of the paper is as follows. Section 2 reviews the economic relevance of tourism demand for the Balearic Islands. In Section 3.1, we introduce some important neural network concepts and introduce the main notation. Section 3.2 discusses parameter estimation and asymptotic theory. The empirical application is presented in Section 4. Finally, Section 5 gives some concluding comments.

2. Tourism demand in the Balearic Islands

In this paper we consider daily air passenger arrivals as a proxy for the daily demand for tourism in the Balearic Islands, Spain. We model and forecast the number of passenger arrivals at three different airports, namely Palma de Mallorca, Ibiza and Mahon, which are located in the islands of Mallorca, Ibiza and Menorca, respectively.

The Balearic Islands has a total population of just over 1 million people and are one of the leading sun and sand destinations in the Mediterranean. During 2006 the Balearic Islands received, by air and by sea, over 12.5 million tourists and, of these, approximately 12 million arrived by plane, and 9.77 million were international tourists. The tourism industry accounts for 48% of the total GDP in the Balearics. However, the tourism industry is affected by seasonality, as it is in many other Mediterranean destinations. Almost 9 million tourists visited the islands between the months of May and September, but only 3.5 million visited during the remaining seven months (CITTIB, 2007). Furthermore, the local economy is not only highly dependent on tourism, but the standardized sun and sand product also predominates, despite the efforts of diversification promoted by public and private initiatives (Aguiló et al., 2005).

The three main islands in the Balearics are Mallorca, Ibiza and Menorca (for purposes of simplicity, data for the small island of Formentera are integrated with Ibiza), and each has an international airport in their respective capital cities of Palma de Mallorca, Ibiza and Mahon. Although all the islands enjoy the same climate, there are differences in their economic structures, the number of tourist arrivals, seasonal patterns, and the profiles of tourists who visit each island. Mallorca accounts for 79% of Balearic regional GDP, while Menorca and Ibiza represent 9% and 12%, respectively (CAIB, 2004). In Mallorca, total demand from tourism corresponds to 34% of the island GDP, in Ibiza this percentage is 44%, and in Menorca tourism demand represents 28% of the island GDP (CAIB, 2004).

In 2006, Mallorca received a total of 9.6 million tourists. Of these, 38.4% were from Germany and 24.2% were from the United Kingdom. In comparison, Ibiza, with 1.87 million visitors, had 35.2% from Britain, 17.1% from Germany and 14.8% from Italy. For Menorca, the British represented 50.3% of tourists, followed by domestic tourism (29.4%) of a total of 1.009 million tourist arrivals in 2006 (CITTIB, 2007). It is worth noting that Menorca and Ibiza experience greater seasonality than does Mallorca. In 2005, 57.8% of the total tourist arrivals in Mallorca stayed during the high season, whereas in Menorca and Ibiza, this figure was as high as 83% (CRE, 2005).

From the details provided above, it is obvious that the tourism industry in the Balearic Islands is critical for the economic well-being of the islands, and is highly dependent on the evolution of air transportation in the region. In fact, the high degree of tourism specialization in the Balearic Islands economy, as compared with other destinations in the Mediterranean, is highly dependent on the creation and development of air routes in the region. Hence, it is reasonable to suggest that the airport is the most important aspect of tourism infrastructure that explains the success of the Balearic Islands as a tourism destination.

Furthermore, the image promoted by each island is different. While Menorca appeals primarily to families, Ibiza attracts a younger market, and Mallorca receives a broader array of tourist segments. As a consequence, the majority of tourists in Menorca enjoy day time activities, the Ibiza visitors are more interested in the night life, while in Mallorca both day and night activities are sought (CITTIB, 2007). These differences suggest that each island should be considered as a different tourism destination for purposes of tourism planning, management and promotion.

Due to the importance of tourism in the Balearics, many researchers have used this destination to analyze different aspects of tourism. In particular, from the demand perspective, Aguiló et al. (2005) and Garin and Montero (2007) estimated price and income elasticities using yearly passenger arrivals data. From a microeconomic perspective, Alegre and Pou (2006) demonstrated the trend of tourists staying for shorter periods. However, it has also been shown that the islands benefit from a high repeat visitation rate (Alegre and Cladera, 2006; Garin and Montero, 2007).

On the supply side, it has been recognized that the islands have reached their maximum carrying capacity, as well as the importance of protecting the natural environment and preserving the local cultural identity (Bujosa and Rossello, 2007; Knowles and Curtis, 1999). The role of tour operators in the commercialization and price structure of the packaged sun and sand product has also been investigated, arriving at the conclusion that British and German tour operators have an oligopolistic position towards accommodation providers and customers (Aguiló et al., 2001). Given this background, we now present an econometric approach to forecast tourism demand in the islands.

3. The proposed model

3.1. Model definition

Let $\mathbf{x}_t = (x_{1t}, \ldots, x_{qt})' \in \mathbb{X} \subseteq \mathbb{R}^q$ be a vector which contains q explanatory variables (covariates or predictor variables) for a continuous univariate response $y_t \in \mathbb{R}$, $t = 1, \ldots, T$. Suppose that the relationship between y_t and $\mathbf{x}_t$ follows a regression model of the form

$$y_t = \mathbb{E}\left[y_t|\mathbf{x}_t\right] + \varepsilon_t = f(\mathbf{x}_t) + \varepsilon_t, \tag{1}$$

where the function $f(\cdot)$ is unknown and, in principle, there are no assumptions about the distribution of the random term ε_t. The idea behind NN models is to approximate the regression model in (1) by

$$y_t = \boldsymbol{\alpha}'\tilde{\mathbf{x}}_t + \sum_{m=1}^{M} \lambda_i F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i) + e_t. \tag{2}$$

The vector $\tilde{\mathbf{x}}_t \in \mathbb{R}^{q+1}$ is defined as $\tilde{\mathbf{x}}_t = [1, \mathbf{x}_t']'$, where $\mathbf{x}_t \in \mathbb{R}^q$ is a vector of lagged values of y_t and/or some exogenous variables. The function $F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i)$, often called the activation function, is the logistic function

$$F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i) = \left[1 + \mathrm{e}^{-(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i)}\right]^{-1}, \tag{3}$$

where $\tilde{\boldsymbol{\omega}}_i = \left[\tilde{\omega}_{1i}, \ldots, \tilde{\omega}_{qi}\right]' \in \mathbb{R}^q$ and $\beta_i \in \mathbb{R}$. The linear combination of these functions in (2) forms the so-called hidden layer. Model (2) with (3) does not contain lags of the error term and is, therefore, called a feedforward NN model. For other choices of the activation function, see Chen et al. (2001).

Most of the recent applied papers concerning NN models have advocated the "black-box" nature of such specifications, claiming that, due to their "universal approximation" capability, NN models are very flexible and are able to describe almost any kind of nonlinear mappings. In fact, NNs may be viewed as a kind of smooth transition regression (van Dijk et al., 2002), where the transition variable is an unknown linear combination of the explanatory variables. In this case, there is an optimal number of hidden units, M, that can be translated as the number of limiting regimes (M is fixed) (see, for example, Trapletti et al. (2000), Medeiros and Veiga (2000, 2005), and Medeiros et al. (2006) for similar interpretations).

On the other hand, when M is large enough, the NN model is a "universal approximator" to any Borel-measurable function over a compact set, and a nonparametric interpretation should be advocated. The number of hidden units increases with the sample size, and NN models can be seen as a sieve-approximator of Grenander (1981). Hornik et al. (1994), Chen and Shen (1998), and Chen and White (1998) provide the technical details (see also Chen et al. (2001)). In both cases, we assume that $\widehat{e_t}$ is a "good" estimator of ε_t.

In this paper we consider the following extension of model (2). We will explicitly consider that $e_t \equiv \varepsilon_t$, making the assumption that the model is correctly specified, that is, $\mathbb{E}(y_t|\mathbf{x}_t) = \boldsymbol{\alpha}'\tilde{\mathbf{x}}_t + \sum_{m=1}^{M} \lambda_i F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i)$.

Definition 1. A parametric model $\mathcal{M}$ is called a neural network regression with GARCH errors, NN–GARCH, if

$$\begin{aligned} y_t &= \boldsymbol{\alpha}'\tilde{\mathbf{x}}_t + \sum_{m=1}^{M} \lambda_i F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i) + \varepsilon_t, \\ \varepsilon_t &= h_t^{1/2} u_t, \\ h_t &= \delta + \phi h_{t-1} + \theta \varepsilon_{t-1}^2, \end{aligned} \tag{4}$$

where the function $F(\cdot)$ is defined as in (3) and $\{u_t\}$ is a sequence of independent and identically distributed zero mean random errors with unit variance, $u_t \sim \mathrm{IID}(0, 1)$.

As it is clear from the model definition, we allow for a GARCH(1,1) process in the conditional variance. We restrict attention to the first-order case as it is the most frequently used in practical applications.

3.2. Model estimation and asymptotic theory

In this section we discuss estimation of the NN–GARCH model and the corresponding asymptotic theory. As the true distribution of u_t is unknown, the parameters of model (4) are estimated by the quasi-maximum likelihood estimator (QMLE). Define the true vector of parameters as $\boldsymbol{\psi}_0 = \left(\boldsymbol{\psi}'_{0,M}, \boldsymbol{\psi}'_{0,V}\right)'$, where the conditional mean parameter vector is defined as $\boldsymbol{\psi}_{0,M} = \left(\boldsymbol{\alpha}', \lambda_1, \ldots, \lambda_M, \tilde{\boldsymbol{\omega}}_1, \ldots, \tilde{\boldsymbol{\omega}}_M, \beta_1, \ldots, \beta_M\right)'$ and $\boldsymbol{\psi}_{0,V} = (\delta, \phi, \theta)'$ is the parameter vector for the conditional variance. Let the quasi-likelihood function evaluated at an arbitrary parameter, $\boldsymbol{\psi}$, be given as

$$\mathcal{L}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T} \ell_t(\boldsymbol{\psi}),$$
$$= \frac{1}{T}\sum_{t=1}^{T}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_t) - \frac{\varepsilon_t^2}{2h_t}\right]. \quad (5)$$

Set $\mathbf{Y}_t = \left(y_t, \mathbf{x}'_t\right)'$. Note that the processes $\mathbf{Y}_t$ and h_t, $t \leq 0$, are unobserved, and hence are only arbitrary constants. Thus, $\mathcal{L}_T(\boldsymbol{\psi})$ is a quasi-log-likelihood function that is not conditional on the true (y_0, h_0), making it suitable for practical applications. However, to prove the asymptotic properties of the QMLE, it is more convenient to work with the unobserved process $\left\{\left(\varepsilon_{u,t}, h_{u,t}\right) : t = 0, \pm 1, \pm 2, \ldots\right\}$.

The unobserved quasi-log-likelihood function, conditional on $\mathcal{F}_0 = (y_0, y_{-1}, y_{-2}, \ldots)$ is

$$\mathcal{L}_{u,T}(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T} \ell_{u,t}(\boldsymbol{\psi})$$
$$= \frac{1}{T}\sum_{t=1}^{T}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(h_{u,t}) - \frac{\varepsilon_{u,t}^2}{2h_{u,t}}\right]. \quad (6)$$

The main difference between $\mathcal{L}_T(\boldsymbol{\psi})$ and $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is that the former is conditional on any initial values, whereas the latter is conditional on an infinite series of past observations. In practical situations, the use of (6) is not possible.

Let

$$\widehat{\boldsymbol{\psi}}_T = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\operatorname{argmax}}\, \mathcal{L}_T(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\operatorname{argmax}}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_t(\boldsymbol{\psi})\right]$$

and

$$\widehat{\boldsymbol{\psi}}_{u,T} = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\operatorname{argmax}}\, \mathcal{L}_{u,T}(\boldsymbol{\psi}) = \underset{\boldsymbol{\psi}\in\boldsymbol{\Psi}}{\operatorname{argmax}}\left[\frac{1}{T}\sum_{t=1}^{T}\ell_{u,t}(\boldsymbol{\psi})\right].$$

Define $\mathcal{L}(\boldsymbol{\psi}) = \mathbb{E}\left[\ell_{u,t}(\boldsymbol{\psi})\right]$. In the following subsections, we discuss the existence of $\mathcal{L}(\boldsymbol{\psi})$ and the identifiability of the NN–GARCH models. Then, in Section 3.2.3, we prove the consistency of $\widehat{\boldsymbol{\psi}}_T$ and $\widehat{\boldsymbol{\psi}}_{u,T}$. We first prove the strong consistency of $\widehat{\boldsymbol{\psi}}_{u,T}$, and then show that

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi}_0)\right| \xrightarrow{a.s.} 0,$$

so that the consistency of $\widehat{\boldsymbol{\psi}}_T$ follows. Asymptotic normality of both estimators is considered in Section 3.2.4. We prove the asymptotic normality of $\widehat{\boldsymbol{\psi}}_{u,T}$. The proof of consistency of $\widehat{\boldsymbol{\psi}}_T$ is straightforward.

3.2.1. Existence of the QMLE

The following theorem proves the existence of $\mathcal{L}(\boldsymbol{\psi})$. It is based on Theorem 2.12 in White (1994), which establishes that, under certain conditions of continuity and measurability of the quasi-loglikelihood function, $\mathcal{L}(\boldsymbol{\psi})$ exists. Consider the following assumption.

Assumption 1. The observed sequence of the real-valued random vector $\{y_t, \mathbf{x}_t\}_{t=1}^{T}$ is a realization of a stationary multivariate stochastic process on a complete probability space generated as in (4), where the sequence $\{\varepsilon_t\}_{t=1}^{T}$ is formed by random variables drawn from an absolutely continuous (with respect to a Lebesgue measure on the real line), positive everywhere distribution, such that $\mathbb{E}[\varepsilon_t] = 0$ and $\mathbb{E}[\varepsilon_t^2] = \sigma^2 < \infty$, $\forall t$. In addition, assume that $\mathbb{E}\left[\varepsilon_t | \mathbf{x}_t, \mathcal{F}_{t-1}\right] = 0$, where $\mathcal{F}_{t-1}$ is the filtration with respect to all past information. Finally, $\mathbb{E}\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = h_t$, where $0 < h_t < \infty, \forall t$.

Theorem 1. *Under Assumption 1, $\mathcal{L}(\boldsymbol{\psi})$ exists and is finite.*

3.2.2. Identifiability of the model

A fundamental problem for statistical inference that haunts nonlinear econometric models is the unidentifiability of the quasi-log-likelihood function. In order to have statistical inference, we need to show that $\boldsymbol{\psi}_0$ is the unique minimizer of $\mathcal{L}(\boldsymbol{\psi})$. However, the unconstrained NN–GARCH model is neither locally nor globally identified. Lack of identification of NN models has been discussed by many authors (Hwang and Ding, 1997; Medeiros et al., 2006). Here, we extend the discussion to the case of the NN–GARCH model. Five properties of the model cause the lack of identifiability:

(P.1) The property of interchangeability of the hidden units. If in (4), a group of functions $F(\tilde{\boldsymbol{\omega}}'_m\mathbf{x}_t - \beta_m)$, $m = 1, \ldots, M$, is permuted, then we can also permute the parameters λ_m yielding the same value of the quasi-likelihood function. This results in several different models that are indistinct among themselves. As a consequence, in the estimation of parameters, we will have several equal local maxima for $\mathcal{L}(\boldsymbol{\psi})$.
(P.2) The fact that $F(\tilde{\boldsymbol{\omega}}'_m\mathbf{x}_t - \beta_m) = 1 - F(-\tilde{\boldsymbol{\omega}}'_m\mathbf{x}_t + \beta_m)$.
(P.3) The presence of irrelevant hidden units. If model (4) has hidden units, such that $\lambda_m = 0$ for at least one m, the parameters $\tilde{\boldsymbol{\omega}}_m$ and β_m remain unidentified. Conversely, if $\tilde{\boldsymbol{\omega}}_m = \mathbf{0}$, then λ_m and β_m can take any value without the value of the likelihood function being affected.
(P.4) No GARCH effects, i.e., $\phi = \theta = 0$. If there are no GARCH effects and we try to estimate an NN–GARCH model, the parameters ϕ and θ are not identifiable.
(P.5) Perfect collinearity among the elements of $\mathbf{x}_t$.

Hence, establishing restrictions on the parameters of (4) that simultaneously avoid any permutation of regimes (property (P.1)), symmetries in the logistic function (property (P.2)), model reducibility (property (P.3)), imposing $\theta > 0$, and avoiding perfect collinerity in $\mathbf{x}_t$, we guarantee the identifiability of the model.

The problem of interchangeability (property (P.1)) can be prevented by considering the following restriction, as in Medeiros et al. (2006):

Restriction 1. *The parameters $\beta_1, \ldots, \beta_M, \lambda_1, \ldots, \lambda_M$ are restricted as follows: $\beta_1 \leq \cdots \leq \beta_M$ or $\lambda_1 \geq \cdots \geq \lambda_M$.*

Again following Medeiros et al. (2006), the consequences due to the symmetry of the logistic function (property (P.2)) and irrelevant hidden units (property (P.3)) can be solved if we impose the following restrictions:

Restriction 2. *The parameters of the NN–GARCH model satisfy the following restrictions:*

(1) $\lambda_m \neq 0, \forall m \in \{1, 2, \ldots, M\}$; *and*
(2) $\tilde{\omega}_{1m} > 0, \forall m \in \{1, 2, \ldots, M\}$.

Restriction 3. *The parameters δ, ϕ, and θ in (4) are such that:*

(1) $\delta > 0, \phi \geq 0, \theta > 0$; *and*
(2) $\phi + \theta < 1$.

Consider now the following assumption.

Assumption 2. The $T \times (q+1)$ matrix

$$\mathbf{X} = \begin{bmatrix} \tilde{\mathbf{x}}_1' \\ \tilde{\mathbf{x}}_2' \\ \vdots \\ \tilde{\mathbf{x}}_T' \end{bmatrix}$$

has full rank, that is, $\mathrm{rank}(\mathbf{X}) = q + 1$, *a.s.* Furthermore, the true parameter vector $\boldsymbol{\psi}_0$ satisfies Restrictions 1–3.

Theorem 2. *Under Assumptions 1 and 2, the NN–GARCH model is globally identifiable. Furthermore, $\mathcal{L}(\boldsymbol{\psi})$ is uniquely maximized at $\boldsymbol{\psi}_0$.*

In practice, the presence of irrelevant regimes can be circumvented by applying a “specific-to-general” model building strategy, as in Medeiros et al. (2006) or regularization techniques as in MacKay (1992a,b). Pruning is also a viable procedure to avoid overparametrization (Anders and Korn, 1999). In the empirical exercise we show the results when the Bayesian regularization approach of MacKay (1992a,b) is adopted.[1]

3.2.3. Consistency

Consider the following assumption.

Assumption 3. The parameter vector $\boldsymbol{\psi}_0 \in \boldsymbol{\Psi}$, which maximizes the quasi-log-likelihood function is in the interior of $\boldsymbol{\Psi}$, a compact subset of a finite dimensional Euclidean space.

Theorem 3. *Under Assumptions 1 and 3, the QMLE $\widehat{\boldsymbol{\psi}}_T$ is strongly consistent for $\boldsymbol{\psi}$, that is, $\widehat{\boldsymbol{\psi}}_T \xrightarrow{a.s.} \boldsymbol{\psi}_0$.*

3.2.4. Asymptotic normality

First, we introduce the following matrices:

$$\mathbf{A}(\boldsymbol{\psi}_0) = \mathbb{E}\left[-\left.\frac{\partial^2 \ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0} \right],$$

$$\mathbf{B}(\boldsymbol{\psi}_0) = \mathbb{E}\left[\left.\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0} \left.\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0} \right]$$

and

$$\begin{aligned} \mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\Bigg[& \frac{1}{2h_t}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial^2 h_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'} \\ & - \frac{1}{2h_t^2}\left(2\frac{\varepsilon_t^2}{h_t} - 1\right)\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'} \\ & + \left(\frac{\varepsilon_t}{h_t^2}\right)\left(\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'} + \frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'}\right) \\ & + \frac{1}{h_t}\left(\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'} + \varepsilon_t\frac{\partial^2\varepsilon_t}{\partial\boldsymbol{\psi}}\right)\Bigg], \end{aligned} \tag{7}$$

$$\begin{aligned} \mathbf{B}_T(\boldsymbol{\psi}) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'} \\ &= \frac{1}{T}\sum_{t=1}^{T}\Bigg[\frac{1}{4h_t^2}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)^2\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'} + \frac{\varepsilon_t^2}{h_t}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'} \\ &\quad - \frac{\varepsilon_t}{2h_t^2}\left(\frac{\varepsilon_t^2}{h_t} - 1\right)\left(\frac{\partial h_t}{\partial\boldsymbol{\psi}}\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}'} + \frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\frac{\partial h_t}{\partial\boldsymbol{\psi}'}\right)\Bigg]. \end{aligned} \tag{8}$$

The following theorem states the asymptotic normality result.

Theorem 4. *Under Assumptions 1–3 and the additional assumption $\mathbb{E}\left[\varepsilon_t^4\right] = \mu_4 < \infty$, then*

$$T^{1/2}(\widehat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0) \xrightarrow{d} \mathrm{N}\left(\mathbf{0}, \boldsymbol{\Omega}_0\right), \tag{9}$$

where $\boldsymbol{\Omega}_0 = \mathbf{A}(\boldsymbol{\psi}_0)^{-1}\mathbf{B}(\boldsymbol{\psi}_0)\mathbf{A}(\boldsymbol{\psi}_0)^{-1}$. Furthermore, the matrices $\mathbf{A}(\boldsymbol{\psi}_0)$ and $\mathbf{B}(\boldsymbol{\psi}_0)$ are consistently estimated by $\mathbf{A}_T(\widehat{\boldsymbol{\psi}}_T)$ and $\mathbf{B}_T(\widehat{\boldsymbol{\psi}}_T)$, respectively.

4. Modelling tourism demand

4.1. Data and experiment

The data set is comprised of daily passenger arrivals at the three international airports in the Balearic Islands, namely Palma de Mallorca, Ibiza and Mahon, which are located on the islands of Mallorca, Ibiza and Menorca, respectively. The data are daily, for the period 1 January 2001 to 31 December 2006, giving a total of 2191 observations. The source of data is the AENA (Aeropuertos Españoles y Navegación Aérea), the Spanish National Airport Authority.

Fig. 1 shows the daily number of air passenger arrivals for the three airports considered in this paper. Several interesting facts emerge from the figure. First, as expected, there is a strong annual cycle with a large increase in the numbers of arrivals during summer. Second, the volatility seems to increase during summer. Although not clear from visual inspection, there is also an obvious strong weekly cycle in all three series. Finally, there is no evidence of a trend (either positive or negative) in any of the three airports. Fig. 2 illustrates the evolution of the annual average number of air passenger arrivals at each airport. Panel (a) in Fig. 2 shows the numbers in levels, while Panel (b) presents the results for logarithms. It is quite clear, especially in logarithms, that the annual average is stable throughout the sample.

Table 1 presents some descriptive statistics. The table shows the mean and variance of daily air passenger arrivals at each airport, with workdays and weekends considered separately for each quarter of the year.

We use data from 2001 to 2005 to estimate the models and leave 2006 to evaluate the forecasting performance of the models. We now describe our model to forecast tourism demand.

4.2. The model

Due to the clear annual and weekly cycle, we consider the following augmented specification of the NN–GARCH model. Set y_t as the logarithm of the daily air passenger arrivals at each airport which is modelled as

$$\begin{aligned} \Delta_7\log(y_t) &= \alpha_0 + \alpha_1\sin\left(\frac{2\pi}{365}t\right) + \alpha_2\cos\left(\frac{2\pi}{365}t\right) \\ &\quad + \boldsymbol{\alpha}_3'\Delta_7\mathbf{y}_{t-1} + \sum_{m=1}^{M}\lambda_m F\left(\tilde{\boldsymbol{\omega}}_m'\Delta_7\mathbf{y}_{t-1} - \beta_m\right) + \varepsilon_t, \end{aligned} \tag{10}$$

[1] Using the sequential strategy of Medeiros et al. (2006) does not substantially change the results.

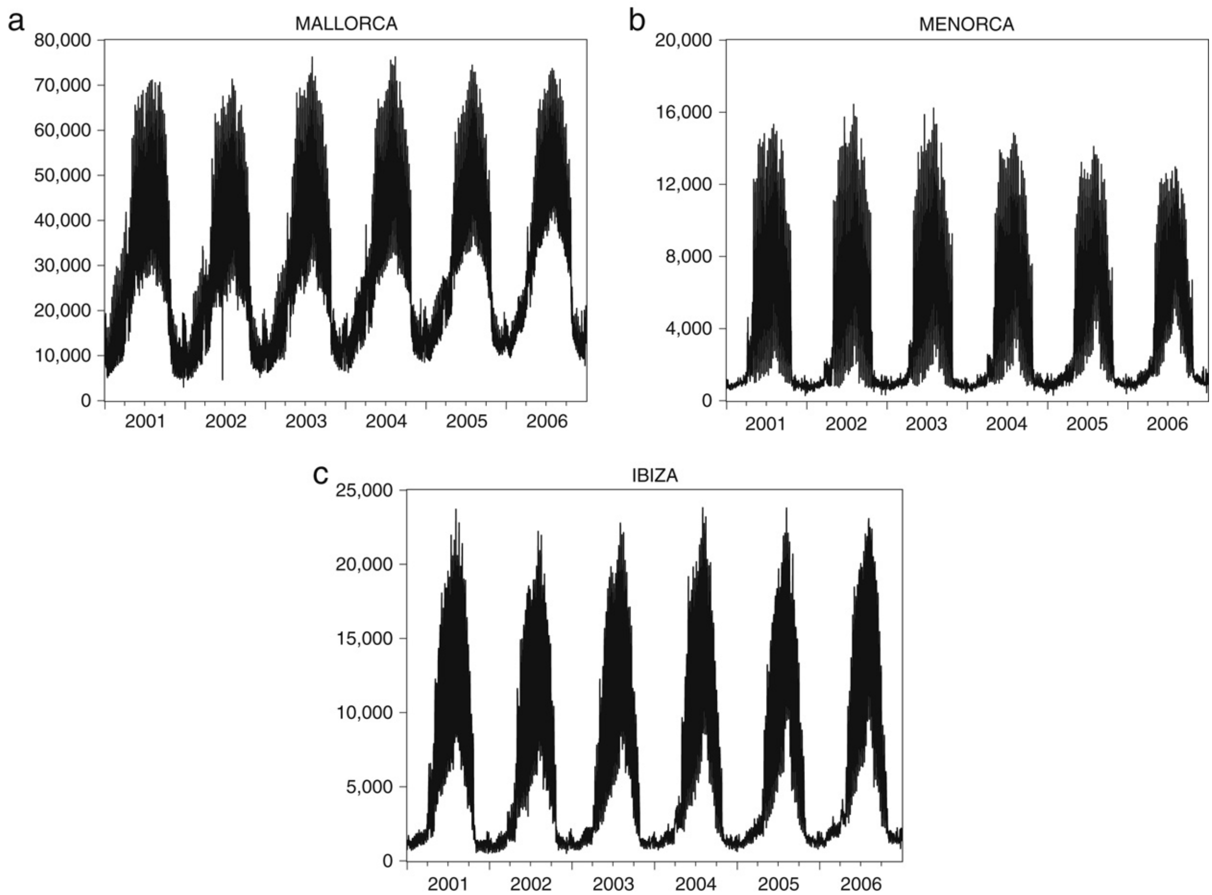

Fig. 1. Daily air passenger arrivals in the three airports of the Balearic Islands, Spain. The data compromise the period from January 1, 2001 to December 31, 2006, a total of 2191 observations. Panel (a) is for Mallorca airport. Panel (b) is for Menorca airport. Panel (c) is for Ibiza airport.

Table 1
Descriptive statistics

Period	Workdays		Weekends		All	
	Mean	Variance	Mean	Variance	Mean	Variance
Mallorca						
Jan–Mar	14,866.99	41,255,055.19	15,601.05	31,188,650.58	15,075.94	38,436,434.55
Apr–Jun	28,295.60	62,922,481.25	46,070.21	187,016,431.23	33,374.06	162,695,027.89
Jul–Sep	36,262.32	50,861,969.62	60,179.93	86,281,122.55	43,108.30	177,946,073.29
Oct–Dec	16,104.42	55,214,479.96	20,816.89	221,803,093.96	17,453.28	107,126,728.71
All	23,911.37	131,303,980.78	35,762.00	465,273,613.36	27,297.27	255,228,758.70
Menorca						
Jan–Mar	1,000.13	81,443.95	899.55	47,973.12	971.50	73,873.61
Apr–Jun	4,409.67	13,827,500.14	5,322.09	7,728,420.79	4,670.36	12,237,740.97
Jul–Sep	6,709.30	15,340,193.93	8,260.81	2,195,921.25	7,153.39	12,059,757.08
Oct–Dec	1,702.37	3,298,598.05	1,770.46	2,225,855.54	1,721.86	2,987,897.11
All	3,463.91	13,302,948.40	4,079.42	11,683,780.87	3,639.77	12,912,135.42
Ibiza						
Jan–Mar	1,574.44	192,996.32	1,277.01	154,448.54	1,489.77	199,763.50
Apr–Jun	5,472.58	8,553,967.02	11,226.13	35,108,318.42	7,116.45	22,858,596.72
Jul–Sep	9,173.44	6,956,277.59	18,612.01	7,536,522.71	11,875.06	25,342,719.00
Oct–Dec	2,159.79	2,467,637.42	3,114.20	10,419,618.57	2,432.97	4,915,407.91
All	4,606.33	13,767,417.84	8,595.33	61,138,621.23	5,746.04	30,529,171.09

The table shows the mean and variance of daily air passenger arrivals at three different airports (Mallorca, Menorca, and Ibiza) in the Balearic islands, Spain. Workdays and weekends are considered separately for each quarter of the year. The period considered ranges from January 1, 2001 to December 31, 2006.

where ε_t follows a GARCH error, such that $\mathbb{E}\left[\varepsilon_t^2|\mathcal{F}_{t-1}\right] \equiv h_t = \delta + \phi h_{t-1} + \theta\varepsilon_{t-1}^2$. $\Delta_7\mathbf{y}_{t-1}$ is a set of lagged values of $\Delta_7 \log(y_t)$, which are determined by some statistical criterion. $\alpha_0, \alpha_1, \alpha_2, \boldsymbol{\alpha}_3, \lambda_1, \ldots, \lambda_M, \tilde{\boldsymbol{\omega}}_1, \ldots, \tilde{\boldsymbol{\omega}}_M, \beta_1, \ldots, \beta_M, \delta, \phi$, and θ are parameters to be estimated. Finally, $\Delta_7 \log(y_t) \equiv \log(y_t) - \log(y_{t-7})$. The seventh-order difference is taken in order to remove the weekly pattern. The seventh-order difference of the logs is also a measure of weekly "returns" on air passenger arrivals. Another possibility is to model y_t directly and to use dummy variables for each day of the week. Both strategies were tested and the one adopted in this

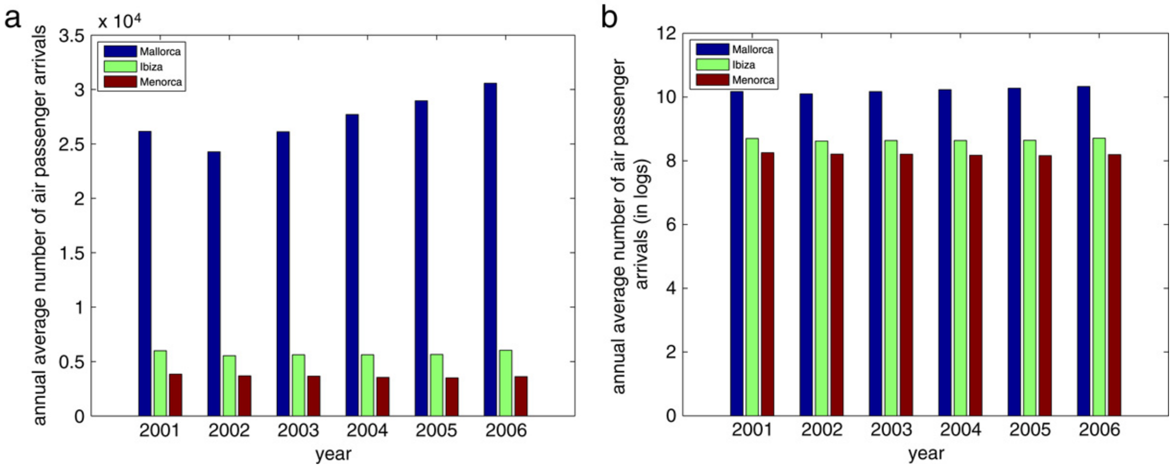

Fig. 2. Annual average number of air passenger arrivals in the three airports of the Balearic Islands, Spain. The data compromise the period from 2001 to 2006. Panel (a) gives the numbers in levels. Panel (b) gives the numbers in logarithms.

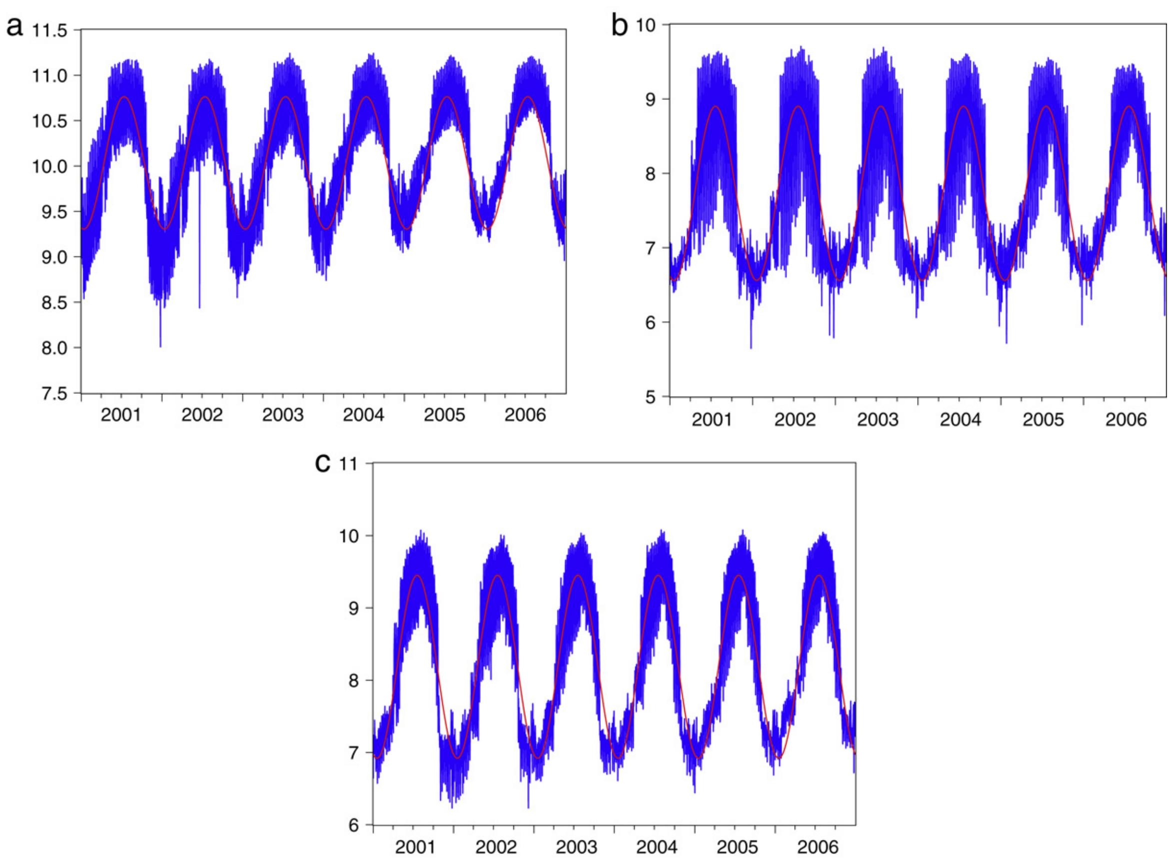

Fig. 3. Daily log air passenger arrivals in the three airports of the Balearic Islands, Spain and fitted harmonic (red line) representing the annual cycle. The data compromise the period from January 1, 2001 to December 31, 2006, a total of 2191 observations. Panel (a) is for Mallorca airport. Panel (b) is for Menorca airport. Panel (c) is for Ibiza airport.

paper was the one with superior in-sample explanatory power. The inclusion of the first two terms on the right-hand side of Eq. (10) is motivated by the strong presence of an annual cycle. Although the presence of such a pattern is not clear in differenced data, the inclusion of such terms reduced the presence of significant higher-order autocorrelations in the residuals. Just for illustration, Fig. 3 shows a fitted harmonic for y_t. We do not consider either deterministic or stochastic trends in the series as it is clear from Fig. 2 that the data (especially in logs) do not have trends. This is also confirmed by standard unit root tests. All the models have been estimated in Matlab and the codes are available from the authors upon written request.

4.3. Results

In this section we report point forecasting results for 2006. The NN–GARCH is compared with a concurrent linear specification and a NN model with no GARCH effects. We compute daily point forecasts from one- to 30-steps-ahead, and aggregate these forecasts in order to obtain monthly forecasts and compare them with the actual numbers.

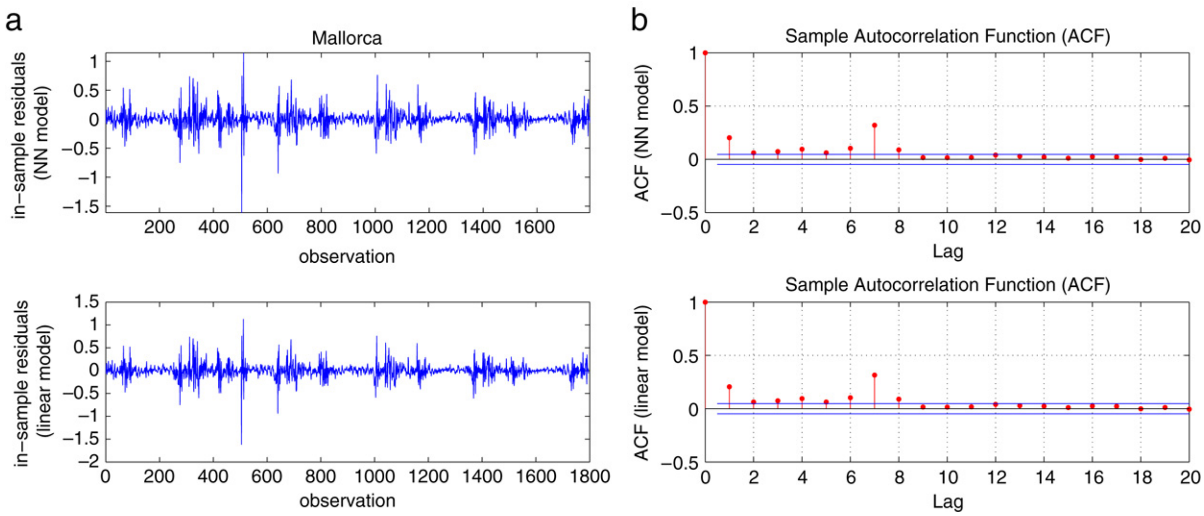

Fig. 4. In-sample residuals (panel (a)) and autocorrelation function for the squared in-sample residuals (panel (b)) for both linear an nonlinear models fitted the series of daily air passenger arrivals at the Mallorca airport.

Table 2
Point forecast results

Model	Workdays		Weekends		All	
	Daily	Monthly	Daily	Monthly	Daily	Monthly
Mallorca						
NN (no GARCH)	5.57	9.63	6.42	9.19	5.95	9.50
NN–GARCH	5.78	9.52	7.38	9.14	6.24	9.41
Linear (no GARCH)	5.60	9.85	6.86	9.46	5.96	9.74
Linear-GARCH	5.66	9.47	6.88	9.11	6.00	9.37
Menorca						
NN (no GARCH)	13.00	15.56	11.94	15.14	12.70	15.44
NN–GARCH	12.22	13.73	11.65	13.30	12.06	13.61
Linear (no GARCH)	12.93	16.71	12.00	16.72	12.66	16.71
Linear-GARCH	12.26	13.84	11.68	13.62	12.00	13.78
Ibiza						
NN (no GARCH)	10.24	17.24	12.95	16.88	11.01	17.14
NN–GARCH	9.93	15.96	13.27	16.35	10.88	16.07
Linear (no GARCH)	10.35	17.52	13.00	16.96	11.10	17.37
Linear-GARCH	10.17	15.73	13.13	16.02	11.01	15.81

The table shows, for each model, the mean absolute percentage errors (MAPE) for the daily one-step-ahead forecasts as well as the MAPE for monthly forecasts based on one to 30-days horizon. Workdays and weekends are considered separately. The dependent variable is the number of air passenger arrivals.

We start by estimating a linear model for each airport without GARCH errors. The number of lags were chosen to minimize the Akaike Information Criterion (AIC) (Akaike, 1973, 1974) among the first 24 lags of y_t. The selected lag orders for Mallorca, Menorca, and Ibiza are 14, 15, and 8, respectively. Neural network models are estimated with the same number of lags as in the linear alternative.[2] We estimated the models with 3, 5, 10, and 15 hidden units. The specification of the NN–GARCH models considered here is based on the Bayesian regularization of MacKay (1992a,b) which controls over-fitting by penalized likelihood. The results for different choices concerning the initial number of hidden units are quite similar and we report only the case for 10 units. As mentioned before, there are other approaches available to control for possible over-parametrization and to avoid the estimation of unidentified NN models. Comparing different methodologies in the present context is certainly a topic for future research.

For both linear and nonlinear models and all three series, there are strong evidences of heteroskedasticity. For example, Fig. 4 shows the in-sample residuals for both linear and nonlinear models fitted to the daily air passenger arrivals at the Mallorca airport as well as the autocorrelation function for the squared in-sample residuals. As can be seen from Fig. 4, the in-sample residuals display volatility clusters and extreme observations. Finally, there are also significant autocorrelations in the squared in-sample squared residuals. We proceed by estimating models with a GARCH(1,1) specification for the conditional variance.

Table 2 depicts the results for point forecasts. Multi-step forecasts are computed by Monte Carlo simulation with the error variance replaced by either the GARCH forecast or the unconditional in-sample residual variance. A total of 500 replications is used in the simulations. When computing multi-step forecasts for the conditional variance, we might get very large values for the predicted conditional variance due the strong persistence in the estimated GARCH models. In these cases, we implement an "insanity filter" and replace the large estimated values by the in-sample unconditional variance. A similar approach was used in Teräsvirta et al. (2005) for the case of conditional mean forecasting.

Analyzing the results, some facts emerge from the table. First, all models have similar performance and the differences are not pronounced apart from few cases. In general, the

[2] The application of the methodology proposed in Medeiros et al. (2006) to select the number of lags in neural network models is also possible.

inclusion of the GARCH term improves the forecasts, specially for monthly forecasts. The only exceptions are the daily one-step-ahead forecasts for the Mallorca airport and the daily forecasts during weekends for the Ibiza airport. When GARCH errors are considered, the performance of the nonlinear model relative to linear alternative is better achieved for Menorca airport. On the other hand, when no GARCH effects are imposed, the neural network model usually outperforms the linear specification.

5. Conclusions

In this paper we developed an alternative approach to analyze the demand for international tourism in the Balearic Islands, Spain, by using a neural network model that incorporated time-varying conditional volatility. This approach to estimating demand is an extension of existing neural network regression models. A detailed analysis of the asymptotic properties of the parameter estimates was presented. Owing to the critical importance of international tourism demand to the economy of the Balearic Islands, accurate forecast tools for gauging international tourism demand is critical. The methods presented have provided a powerful tool to policymakers and tourism management officials alike in the Balearic Islands and other tourism intensive economies. Moreover, time-varying variances provide useful information regarding the risks associated with variations in international tourist arrivals.

Acknowledgments

The first author is thankful to CNPq/Brazil for partial financial support. The second author wishes to acknowledge the financial support of the Australian Research Council, the Japanese Ministry of Education, and the Faculty of Economics at Yokohama National University. The suggestions of two anonymous referees are greatly acknowledged. The last two authors acknowledge the financial support from the Government of the Balearic Islands through the "Dicecció General d'Investigació, Desenvolupament Tecnológic i Innovació" PROGECIB-14B and AENA for providing the data used in the paper.

Appendix A. Proofs

Define $\mathbf{w}_t = (y_t, h_t, u_t)'$ and note that

$$\mathbf{w}_t = \mathbf{P}\left(\mathbf{w}_{t-1}, \mathbf{x}_t\right) + \boldsymbol{\eta}_t, \tag{A.1}$$

where

$$\mathbf{P}\left(\mathbf{w}_{t-1}, \mathbf{x}_t\right) = \begin{pmatrix} \boldsymbol{\alpha}'\tilde{\mathbf{x}}_t + \sum_{m=1}^{M} \lambda_i F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i) \\ \delta + \left(\phi + \theta u_{t-1}^2\right) h_{t-1} \\ 0 \end{pmatrix}$$

and $\boldsymbol{\eta}_t = (\varepsilon_t, 0, u_t)'$.

Proof of Theorem 1. It is easy to see that $\mathbf{P}(\mathbf{w}_{t-1}, \mathbf{x}_t)$, as in (A.1), is a continuous function in the parameter vector $\boldsymbol{\psi}$. Similarly, we can see that $\mathbf{P}(\mathbf{w}_{t-1}, \mathbf{x}_t)$ is continuous in $\mathbf{x}_t$, and therefore is measurable, for each fixed value of $\boldsymbol{\psi}$.

Furthermore, under the Assumption 1 and the restrictions in Assumption 2, then $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|h_{u,t}\right|\right] < \infty$ and $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|y_{u,t}\right|\right] < \infty$. By Jensen's inequality, $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\ln \left|h_{u,t}\right|\right|\right] < \infty$. Thus, $\mathbb{E}\left[\left|\ell_{u,t}(\boldsymbol{\psi})\right|\right] < \infty \,\forall\, \boldsymbol{\psi} \in \boldsymbol{\Psi}$. ■

Proof of Theorem 2. Following the same lines as in Hwang and Ding (1997) and Suarez-Fariñas et al. (2004) it is trivial to show that the conditions in Assumption 2 guarantees that the NN–GARCH has no two equivalent observable structures.

Let $h_{0,t}$ be the true conditional variance and $\varepsilon_{0,t} = h_{0,t}^{1/2} u_t$. In order to show that $\mathcal{L}(\boldsymbol{\psi})$ is uniquely maximized at $\boldsymbol{\psi}_0$, rewrite the maximization problem as

$$\max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left[\mathcal{L}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi}_0)\right] = \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\{\mathbb{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{\varepsilon_t^2}{h_{u,t}} + 1\right]\right\}. \tag{A.2}$$

Writing $\varepsilon_t = \varepsilon_t - \varepsilon_{0,t} + \varepsilon_{0,t}$, Eq. (A.2) becomes

$$\begin{aligned} \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left[\mathcal{L}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi}_0)\right] &= \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\{\mathbb{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{h_{0,t}}{h_{u,t}} + 1\right]\right. \\ &\quad \left. - \mathbb{E}\left[\frac{\left[\varepsilon_t - \varepsilon_{0,t}\right]^2}{h_{u,t}}\right] - \mathbb{E}\left[\frac{2\eta_t h_{0,t}^{1/2}\left(\varepsilon_t - \varepsilon_{0,t}\right)}{h_{u,t}}\right]\right\} \\ &= \max_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left\{\mathbb{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{h_{0,t}}{h_{u,t}} + 1\right] - \mathbb{E}\left[\frac{\left[\varepsilon_t - \varepsilon_{0,t}\right]^2}{h_{u,t}}\right]\right\}, \end{aligned} \tag{A.3}$$

where

$$\mathbb{E}\left[\frac{2u_t h_{0,t}^{1/2}\left(\varepsilon_t - \varepsilon_{0,t}\right)}{h_{u,t}}\right] = 0$$

by the Law of Iterated Expectations.

Note that, for any $x > 0$, $m(x) = \ln(x) - x \leq 0$, so that

$$\mathbb{E}\left[\ln\left(\frac{h_{0,t}}{h_{u,t}}\right) - \frac{h_{0,t}}{h_{u,t}}\right] \leq 0.$$

Furthermore, $m(x)$ is maximized at $x = 1$. If $x \neq 1$, $m(x) < m(1)$, implying that $\mathbb{E}[m(x)] \leq \mathbb{E}[m(1)]$, with equality only if $x = 1$ a.s. However, this will occur only if $\frac{h_{0,t}}{h_{u,t}} = 1$, a.s. In addition,

$$\mathbb{E}\left[\frac{\left[\varepsilon_t - \varepsilon_{0,t}\right]^2}{h_{u,t}}\right] = 0$$

if and only if $\varepsilon_t = \varepsilon_{0,t}$. Hence, $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. This completes the proof. ■

Proof of Theorem 3. Following White (1994), Theorem 3.5, $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{a.s.} \boldsymbol{\psi}_0$ if the following conditions hold:

(1) The parameter space $\boldsymbol{\Psi}$ is compact.
(2) $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is continuous in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. Furthermore, $\mathcal{L}_{u,T}(\boldsymbol{\psi})$ is a measurable function of y_t, $t = 1, \ldots, T$, for all $\boldsymbol{\psi} \in \boldsymbol{\Psi}$.
(3) $\mathcal{L}(\boldsymbol{\psi})$ has a unique maximum at $\boldsymbol{\psi}_0$.
(4) $\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi})\right| = 0$, *a.s.*

Condition (1) holds by assumption. Theorem 1 shows that Conditions (2) and (3) are satisfied. By Lemma 1, Condition (4) is also satisfied. Thus, $\widehat{\boldsymbol{\psi}}_{u,T} \xrightarrow{a.s.} \boldsymbol{\psi}_0$.

Lemma 2 shows that

$$\lim_{T\to\infty} \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})\right| = 0 \quad a.s.,$$

implying that $\widehat{\boldsymbol{\psi}}_T \xrightarrow{a.s.} \boldsymbol{\psi}_0$. This completes the proof. ■

Proof of Theorem 4. We start by proving asymptotic normality of the QMLE using the unobserved log-likelihood. When this is shown, the proof using the observed log-likelihood is immediate by Lemmas 2 and 4. According to Theorem 6.4 in White (1994), to prove the asymptotic normality of the QMLE we need the following conditions in addition to those stated in the proof of Theorem 3:

(5) The true parameter vector $\boldsymbol{\psi}_0$ is interior to $\boldsymbol{\Psi}$.
(6) The matrix

$$\mathbf{A}_T(\boldsymbol{\psi}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 \ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right)$$

exists *a.s.* and is continuous in $\boldsymbol{\Psi}$.
(7) The matrix $\mathbf{A}_T(\boldsymbol{\psi}) \xrightarrow{a.s.} \mathbf{A}(\boldsymbol{\psi}_0)$, for any sequence $\boldsymbol{\psi}_T$, such that $\boldsymbol{\psi}_T \xrightarrow{a.s.} \boldsymbol{\psi}_0$.
(8) The score vector satisfies

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{B}(\boldsymbol{\psi}_0)).$$

Condition (5) is satisfied by assumption. Condition (6) follows from the fact that $\ell_t(\boldsymbol{\psi})$ is differentiable of order two on $\boldsymbol{\psi} \in \boldsymbol{\Psi}$, and the stationarity of the NN–GARCH model. The non-singularity of $\mathbf{A}(\boldsymbol{\psi}_0)$ and $\mathbf{B}(\boldsymbol{\psi}_0)$ follows from Lemma 4. Furthermore, Lemmas 3 and 5 imply that Condition (7) is satisfied. In Lemma 6, we prove that condition (8) is also satisfied. This completes the proof. ■

Appendix B. Lemmas

Lemma 1. *Suppose that y_t follows an NN–GARCH model satisfying the restrictions in Assumptions 1 and 2, and the stationarity and ergodicity conditions are met. Then,*

$$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}(\boldsymbol{\psi})\right| = 0, \quad a.s.$$

Proof. Set $g(\mathbf{Y}_t, \boldsymbol{\psi}) = \ell_{u,t}(\boldsymbol{\psi}) - \mathbb{E}\left[\ell_{u,t}(\boldsymbol{\psi})\right]$, where $\mathbf{Y}_t = \left[y_t, y_{t-1}, y_{t-2}, \ldots, \mathbf{x}_t', \mathbf{x}_{t-1}', \ldots\right]'$. Hence, $\mathbb{E}\left[g(\mathbf{Y}_t, \boldsymbol{\psi})\right] = 0$. It is clear that $\mathbb{E}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}|g(\mathbf{Y}_t, \boldsymbol{\psi})|\right] < \infty$ by Theorem 1 and the stationarity assumption. Furthermore, as $g(\mathbf{Y}_t, \boldsymbol{\psi})$ is strictly stationary and ergodic, then $\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|T^{-1}\sum_{t=1}^{T} g(\mathbf{Y}_t, \boldsymbol{\psi})\right| = 0$, *a.s.*. This completes the proof. ■

Lemma 2. *Under the assumptions of Lemma 1,*

$$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})\right| = 0, a.s.$$

Proof. First, write

$$h_t = \sum_{i=0}^{t-1}\phi^i\left(\delta + \theta\varepsilon_{t-1-i}^2\right) + \phi^t h_0$$

and

$$h_{u,t} = \phi^{t-1}\left(\delta + \theta\varepsilon_{u,0}^2\right) + \sum_{i=0}^{t-2}\phi^i\left(\delta + \theta\varepsilon_{t-1-i}^2\right) + \phi^t h_{u,0},$$

such that

$$\begin{aligned}\left|h_t - h_{u,t}\right| &= |\phi^{t-1}\theta\left(\varepsilon_0^2 - \varepsilon_{u,0}^2\right) + \phi^t\left(h_0 - h_{u,0}\right)| \\ &\le \phi^{t-1}\theta\left|\varepsilon_0^2 - \varepsilon_{u,0}^2\right| + \phi^t\left|h_0 - h_{u,0}\right|.\end{aligned}$$

Under the stationarity of the process, it is clear that $0 < \phi < 1$. Furthermore, $h_{u,0}$ and $\varepsilon_{0,u}^2$ are well defined, as

$$\mathbb{P}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left(h_{u,0} > K_1\right)\right] \to 0 \quad \text{as } K_1 \to \infty, \quad \text{and}$$

$$\mathbb{P}\left[\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left(\varepsilon_{u,0}^2 > K_2\right)\right] \to 0 \quad \text{as } K_2 \to \infty.$$

Thus,

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|h_t - h_{u,t}\right| \le K_h\rho_1^t, \quad a.s.$$

and

$$\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\varepsilon_0^2 - \varepsilon_{u,0}^2\right| \le K_\varepsilon\rho_2^t, \quad a.s.,$$

where K_h and K_ε are positive and finite constants, $0 < \rho_1 < 1$, and $0 < \rho_2 < 1$. Hence, as $h_t > \delta$ and $\log(x) \le x - 1$,

$$\begin{aligned}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\ell_t - \ell_{u,t}\right| &\le \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left[\varepsilon_t^2\left|\frac{h_{u,t} - h_t}{h_t h_{u,t}}\right| + \left|\log\left(1 + \frac{h_t - h_{u,t}}{h_{u,t}}\right)\right|\right] \\ &\le \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left(\frac{1}{\delta^2}\right)K_h\rho_1^t\varepsilon_t^2 + \sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left(\frac{1}{\delta}\right)K_h\rho_1^t, \quad a.s.\end{aligned}$$

Following the same arguments as in the proofs of Theorems 2.1 and 3.1 in Francq and Zakoïan (2004), it can be shown that

$$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left|\mathcal{L}_{u,T}(\boldsymbol{\psi}) - \mathcal{L}_T(\boldsymbol{\psi})\right| = 0, a.s.$$

This completes the proof. ■

Lemma 3. *Under the conditions of Theorem 4,*

$$\mathbb{E}\left[\left|\left.\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\right|\right] < \infty, \tag{B.4}$$

$$\mathbb{E}\left[\left|\left.\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\left.\frac{\partial\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right|\right] < \infty, \quad \textit{and} \tag{B.5}$$

$$\mathbb{E}\left[\left|\left.\frac{\partial^2\ell_t(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}\right|\right] < \infty. \tag{B.6}$$

Proof. Set

$$\nabla_0\ell_{u,t} \equiv \left.\frac{\partial\ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}, \qquad \nabla_0 h_{u,t} \equiv \left.\frac{\partial h_{u,t}}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0},$$

$$\nabla_0\varepsilon_t \equiv \left.\frac{\partial\varepsilon_t}{\partial\boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0},$$

$$\nabla_0^2\ell_{u,t} \equiv \left.\frac{\partial^2\ell_{u,t}(\boldsymbol{\psi})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}, \qquad \nabla_0^2 h_{u,t} \equiv \left.\frac{\partial^2 h_{u,t}}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}, \quad \text{and}$$

$$\nabla_0^2\varepsilon_t \equiv \left.\frac{\partial^2\varepsilon_t}{\partial\boldsymbol{\psi}\partial\boldsymbol{\psi}'}\right|_{\boldsymbol{\psi}_0}.$$

Then,

$$\nabla_0\ell_{u,t} = \frac{1}{2h_{u,t}}\left(\frac{\varepsilon_t^2}{h_{u,t}} - 1\right)\nabla_0 h_{u,t} - \frac{\varepsilon_t}{h_{u,t}}\nabla_0\varepsilon_t$$

and

$$\begin{aligned}\nabla_0^2\ell_{u,t} &= \left(\frac{\varepsilon_t^2}{h_{u,t}} - 1\right)\frac{1}{2h_{u,t}}\nabla_0^2 h_{u,t} \\ &\quad - \frac{1}{2h_{u,t}^2}\left(2\frac{\varepsilon_t^2}{h_{u,t}} - 1\right)\nabla_0 h_{u,t}\nabla_0 h_{u,t}' \\ &\quad + \left(\frac{\varepsilon_t}{h_{u,t}^2}\right)\left(\nabla_0\varepsilon_t\nabla_0 h_{u,t}' + \nabla_0 h_{u,t}\nabla_0\varepsilon_t'\right) \\ &\quad + \frac{1}{h_{u,t}}\left(\nabla_0\varepsilon_t\nabla_0\varepsilon_t' + \varepsilon_t\nabla_0^2\varepsilon_t\right).\end{aligned}$$

Set $\boldsymbol{\psi} = \left(\boldsymbol{\psi}_M', \boldsymbol{\psi}_V'\right)'$, where $\boldsymbol{\psi}_M$ is the vector of parameters of the conditional mean and $\boldsymbol{\psi}_V$ is the vector of parameters of the conditional variance. As in the proof of Theorem 3.2 in Francq and Zakoïan (2004), the derivatives with respect to $\boldsymbol{\psi}_V$ are clearly bounded. We proceed by analyzing the derivatives with respect to $\boldsymbol{\psi}_M$. Define $g(\mathbf{x}_t;\boldsymbol{\psi}_M) \equiv \boldsymbol{\alpha}'\tilde{\mathbf{x}}_t + \sum_{m=1}^{M}\lambda_i F(\tilde{\boldsymbol{\omega}}_i'\mathbf{x}_t - \beta_i)$. As $\varepsilon_t = y_t - g(\mathbf{x}_t;\boldsymbol{\psi}_M)$, we have

$$\frac{\partial \varepsilon_t}{\partial \boldsymbol{\psi}_M} = -\frac{\partial g(\mathbf{x}_t,\boldsymbol{\psi}_M)}{\partial \boldsymbol{\psi}_M}, \tag{B.7}$$

$$\frac{\partial^2 \varepsilon_t}{\partial \boldsymbol{\psi}_M \partial \boldsymbol{\psi}_M'} = -\frac{\partial^2 g(\mathbf{x}_t,\boldsymbol{\psi}_M)}{\partial \boldsymbol{\psi}_M \partial \boldsymbol{\psi}_M'}, \tag{B.8}$$

$$\frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}_M} = 2\theta \sum_{i=0}^{\infty}\left(\phi^i \varepsilon_{t-1-i}\frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\psi}_M}\right), \quad \text{and} \tag{B.9}$$

$$\frac{\partial^2 h_{u,t}}{\partial \boldsymbol{\psi}_M \partial \boldsymbol{\psi}_M'} = 2\theta \sum_{i=0}^{\infty}\phi^i\left(\varepsilon_{t-1-i}\frac{\partial^2 \varepsilon_{t-1-i}}{\partial \boldsymbol{\psi}_M \partial \boldsymbol{\psi}_M'} + \frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\psi}_M}\frac{\partial \varepsilon_{t-1-i}}{\partial \boldsymbol{\psi}_M'}\right). \tag{B.10}$$

As the derivatives of the transition function are bounded, if stationarity holds, (B.7)–(B.10) are clearly bounded. Hence, the remainder of the proof follows from the proof of Theorem 3.2 (part (i)) in Francq and Zakoïan (2004). This completes the proof. ■

Lemma 4. *Under the conditions of Theorem 4, $\mathbf{A}(\boldsymbol{\psi}_0)$ and $\mathbf{B}(\boldsymbol{\psi}_0)$ are nonsingular and, when u_t has a symmetric distribution, are block-diagonal.*

Proof. First, note that the restrictions in Assumption 2 guarantee the minimality (identifiability) of the NN–GARCH model considered in this paper. Therefore, the results follow from the proof of Theorem 3.2 (part (ii)) in Francq and Zakoïan (2004). This completes the proof. ■

Lemma 5. *Under the conditions of Theorem 4,*

(a) $$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} - \frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right]\right\| = \mathbf{0}, \quad a.s.,$$

(b) $$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial^2 \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'} - \frac{\partial^2 \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right]\right\| = \mathbf{0}, \quad a.s, \quad \text{and}$$

(c) $$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\|\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2 \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'} - \mathbb{E}\left[\frac{\partial^2 \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'}\right]\right\| = \mathbf{0}, \quad a.s.$$

Proof. First, assume that h_0 and $h_{u,0}$ are fixed constants. It is easy to show that

$$\begin{aligned}\left|\frac{\partial h_t}{\partial \boldsymbol{\psi}_M} - \frac{\partial h_{u,t}}{\partial \boldsymbol{\psi}_M}\right| &= 2\theta\phi^{t-1}\left|\varepsilon_0\frac{\partial \varepsilon_0}{\partial \boldsymbol{\psi}_M} - \varepsilon_{u,0}\frac{\partial \varepsilon_{u,0}}{\partial \boldsymbol{\psi}_M}\right| \\ &\leq 2\theta\phi^{t-1}\left(\left|\varepsilon_0\frac{\partial \varepsilon_0}{\partial \boldsymbol{\psi}_M}\right| + \left|\varepsilon_{u,0}\frac{\partial \varepsilon_{u,0}}{\partial \boldsymbol{\psi}_M}\right|\right) < \infty,\end{aligned}$$

as $0 < \phi < 1$ and y_t is stationary and ergodic. Hence, following the same arguments as in the proof of Theorem 3.2 (part (iii)) in Francq and Zakoïan (2004), it is straightforward to show that

$$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}_M} - \frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}_M}\right]\right\| = \mathbf{0}.$$

Furthermore, as

$$\begin{aligned}&\frac{\partial h_t}{\partial \delta} - \frac{\partial h_{u,t}}{\partial \delta} = 0,\\ &\frac{\partial h_t}{\partial \theta} - \frac{\partial h_{u,t}}{\partial \theta} = \varepsilon_0^2 - \varepsilon_{u,0}^2, \quad \text{and}\\ &\frac{\partial h_t}{\partial \phi} - \frac{\partial h_{u,t}}{\partial \phi} = (t-1)\phi^{t-2}\left(\varepsilon_0^2 - \varepsilon_{u,0}^2\right) + t\phi^{t-1}\left(h_0 - h_{u,0}\right),\end{aligned}$$

it is clear that

$$\lim_{T\to\infty}\sup_{\boldsymbol{\psi}\in\boldsymbol{\Psi}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left[\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}_V} - \frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}_V}\right]\right\| = \mathbf{0}.$$

The proof of part (a) is now complete. The proof of part (b) follows along similar lines. The proof of part (c) follows the same arguments as in the proof of Theorem 3.2 (part (v)) in Francq and Zakoïan (2004). This completes the proof. ■

Lemma 6. *Under the conditions of Theorem 4,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0} \xrightarrow{d} \mathrm{N}(\mathbf{0},\mathbf{B}(\boldsymbol{\psi}_0)).$$

Proof. Let $S_T = \sum_{t=1}^{T}\mathbf{c}'\nabla_0\ell_{u,t}$, where $\mathbf{c}$ is a constant vector. Then S_T is a martingale with respect to $\mathcal{F}_t$, the filtration generated by all past observations of y_t. By the given assumptions, $\mathbb{E}[S_T] > 0$. Using the central limit theorem of Stout (1974),

$$T^{-1/2}S_T \xrightarrow{d} \mathrm{N}\left(0, \mathbf{c}'\mathbf{B}(\boldsymbol{\psi}_0)\mathbf{c}\right).$$

By the Cramér–Wold device,

$$T^{-1/2}\sum_{t=1}^{T}\left.\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0} \xrightarrow{d} \mathrm{N}\left(0, \mathbf{B}(\boldsymbol{\psi}_0)\right).$$

By Lemma 5,

$$T^{-1/2}\sum_{t=1}^{T}\left\|\left.\frac{\partial \ell_{u,t}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0} - \left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0}\right\| \xrightarrow{a.s.} \mathbf{0}.$$

Thus,

$$T^{-1/2}\sum_{t=1}^{T}\left.\frac{\partial \ell_t(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}\right|_{\boldsymbol{\psi}_0} \xrightarrow{d} \mathrm{N}(0,\mathbf{B}_0).$$

This completes the proof. ■

References

Aguiló, E., Alegre, J., Riera, A., 2001. Determinants of the price of German tourist packages on the island of Mallorca. Tourism Economics 7, 59–74.

Aguiló, E., Riera, A., Rosselló, J., 2005. The short-term price effect of a tourist tax through a dynamic demand model: The case of the Balearic Islands. Tourism Economics 26, 359–365.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Czáki, F. (Eds.), Second International Symposium on Information Theory. Budapest. Akademiai Kiadó, pp. 267–281.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 19, 716–723.

Alegre, J., Cladera, M., 2006. Repeat visitation in mature sun and sand holiday destinations. Journal of Travel Research 44, 288–297.

Alegre, J., Pou, L., 2006. The length of stay in the demand for tourism. Tourism Management 27, 1343–1355.

Anders, U., Korn, O., 1999. Model selection in neural networks. Neural Networks 12, 309–323.

Andersen, T., Bollerslev, T., Diebold, F., 2006. Parametric and nonparametric measurement of volatility. In: Elliott, G., Granger, C., Timmermann, A. (Eds.), Handbook of Financial Econometrics. North-Holland, Amsterdam.

Bartolomé, A., McAleer, M., Ramos, V., Rey-Maquieira, J., 2008. Risk Management for Air Passenger and International Tourist Arrivals in the Balearic Islands, Spain. Discussion Paper. Faculty of Economics and Business, University of the Balearic Islands.

Basmann, R., Molina, D., Slottje, D., 1987. Price-dependent preferences and the Fechner–Thurstone direct utility function. Journal of Institutional and Theoretical Economics 143, 568–594.

Basmann, R., Slottje, D., 1987. The sensitivity of true-cost-of-living indexes to price-and income-induced changes in aggregate consumers' tastes. Journal of Business and Economic Statistics 5, 483–498.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 21, 307–328.

Bujosa, A., Rossello, J., 2007. Modelling environmental attitudes towards tourism. Tourism Management 28, 688–695.

CAIB, 2004. Input/Output. Marco de las Islas Baleares 2004. Discussion Paper. Palma de Mallorca: Govern de les Illes Balears.
Chan, F., Lim, C., McAleer, M., 2005. Modelling multivariate international tourism demand and volatility. Tourism Management 26, 459–471.
Chan, K.S., Tong, H., 1986. On estimating thresholds in autoregressive models. Journal of Time Series Analysis 7, 179–190.
Chen, X., Racine, J., Swanson, N.R., 2001. Semiparametric ARX neural-network models with an application to forecasting inflation. IEEE Transactions on Neural Networks 12, 674–683.
Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66, 289–314.
Chen, X., White, H., 1998. Improved rates and asymptotic normality for nonparametric neural network estimators. IEEE Transactions on Information Theory 18, 17–39.
CITTIB, 2007. El turisme de les Illes Balears. Dades informatives 2006. Discussion Paper. Palma de Mallorca. Spain: Centre d'Investigació i de Tecnologies Turístiques.
CRE, 2005. Informe econòmic i social de les Illes Balears. Discussion Paper. Palma de Mallorca: Centre de Recerca Econòmica.
Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of UK inflation. Econometrica 50, 987–1007.
Fan, J., Yao, Q., 2003. Nonlinear Time Series: Nonparametric and Parametric Methods. Springer-Verlag, New York, NY.
Francq, C., Zakoïan, J.-M., 2004. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli 10, 605–637.
Gallant, A.R., White, H., 1992. On learning the derivatives of an unknown mapping with multilayer feedforward networks. Neural Networks 5, 129–138.
Garin, T., Montero, L., 2007. Tourism in the Balearic Islands: A dynamic model for international demand using panel data. Tourism Management 28, 1224–1235.
Granger, C.W.J., Teräsvirta, T., 1993. Modelling Nonlinear Economic Relationships. Oxford University Press, Oxford.
Grenander, U., 1981. Abstract Inference. Wiley Series, New York.
Hornik, K., Stinchcombe, M., White, H., Auer, P., 1994. Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives. Neural Computation 6, 1262–1274.
Hornik, K., Stinchombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multi-layer feedforward networks. Neural Networks 3, 551–560.
Hoti, S., McAleer, M., Shareef, R., 2007. Modelling international tourism and country risk spillovers for Cyprus and Malta. Tourism Management 28, 1472–1484.
Hwang, J.T.G., Ding, A.A., 1997. Prediction intervals for artificial neural networks. Journal of the American Statistical Association 92, 109–125.
Härdle, W., 1990. Applied Nonparametric Regression. Cambridge University Press, Cambridge.
Härdle, W., Lütkepohl, H., Chen, R., 1997. A review of nonparametric time series analysis. International Statistical Review 65, 49–72.
Knowles, T., Curtis, S., 1999. The market viability of European mass tourist destinations. A post-stagnation life-cycle analysis. International Journal of Tourism Research 1, 87–96.
Kuan, C.M., White, H., 1994. Artificial neural networks: An econometric perspective. Econometric Reviews 13, 1–91.
Luukkonen, R., Saikkonen, P., Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. Biometrika 75, 491–499.
MacKay, D.J.C., 1992a. Bayesian interpolation. Neural Computation 4, 415–447.
MacKay, D.J.C., 1992b. A practical Bayesian framework for Backpropagation Networks. Neural Computation 4, 448–472.
McAleer, M., 2005. Automated inference and learning in modeling financial volatility. Econometric Theory 21, 232–261.
Medeiros, M., Teräsvirta, T., Rech, G., 2006. Building neural network models for time series: A statistical approach. Journal of Forecasting 25, 49–75.
Medeiros, M.C., Veiga, A., 2000. A hybrid linear-neural model for time series forecasting. IEEE Transactions on Neural Networks 11, 1402–1412.
Medeiros, M.C., Veiga, A., 2005. A flexible coefficient smooth transition time series model. IEEE Transactions on Neural Networks 16, 97–113.
Poon, S., Granger, C., 2003. Forecasting volatility in financial markets. Journal of Economic Literature 41, 478–539.
Samuelson, P., 1947. Foundations of Economic Analysis. In: Harvard Economic Studies.
Shareef, R., McAleer, M., 2007. Modelling the uncertainty in monthly international tourist arrivals to the Maldives. Tourism Management 28, 23–45.
Stout, W.F., 1974. Almost Sure Convergence. Academic Press, New York.
Suarez-Fariñas, M., Pedreira, C.E., Medeiros, M.C., 2004. Local global neural networks: A new approach for nonlinear time series modeling. Journal of the American Statistical Association 99, 1092–1107.
Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. Journal of the American Statistical Association 89, 208–218.
Teräsvirta, T., van Dijk, D., Medeiros, M., 2005. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. International Journal of Forecasting 21, 755–774.
Theil, H., 1975. Theory and Measurement of Consumer Demand. North-Holland, Amsterdam.
Tong, H., 1978. On a threshold model. In: Chen, C.H. (Ed.), Pattern Recognition and Signal Processing. Sijthoff and Noordhoff, Amsterdam.
Tong, H., 1990. Non-linear Time Series: A Dynamical Systems Approach. Oxford University Press, Oxford.
Tong, H., Lim, K., 1980. Threshold autoregression, limit cycles and cyclical data (with discussion). Journal of the Royal Statistical Society, Series B 42, 245–292.
Trapletti, A., Leisch, F., Hornik, K., 2000. Stationary and integrated autoregressive neural network processes. Neural Computation 12, 2427–2450.
van Dijk, D., Teräsvirta, T., Franses, P.H., 2002. Smooth transition autoregressive models — A survey of recent developments. Econometric Reviews 21, 1–47.
White, H., 1994. Estimation, inference and specification analysis. Cambridge University Press, New York, NY.