Forecasting Tourism Growth with State-Dependent Models

Abstract

We introduce two forecasting methods based on a general class of non-linear models called 'State-Dependent Models' (SDMs) for tourism demand forecasting. Using a Monte Carlo simulation which generated data from linear and non-linear models, we evidence how estimations from SDMs can capture the level shifts pattern and nonlinearity in data. Next, we apply two new forecasting methods based on SDMs to forecast tourism demand growth in Japan. The forecasts are compared with classical recursive SDM forecasting, Naïve forecasting, ARIMA, Exponential Smoothing, Neural Network models, Time varying parameters, Smooth Transition Autoregressive models, and with a linear regression model with two dummy variables. We find that improvements in forecasting with the proposed SDM-based forecasting methods are more pronounced in the longer-term horizons.

Keywords: State-Dependent Models; tourism demand; forecasting; non-linear; Japan.

1. Introduction

Tourism is not only a significant contributor to the local gross domestic product in many places but also the most vulnerable to crisis events (Zhang et al. 2021). This vulnerability and the correlation of tourism data with different types of volatility and business cycles makes forecasting tourism demand a complex exercise (Song et al. 2019). Whilst the ongoing pandemic has disrupted travel and tourism, the World Travel & Tourism Council (2020) predicted a strong summer of travel ahead with a significant rise in forward bookings reported by major travel companies. Therefore, modelling, estimating, and generating accurate forecasts of tourism demand continues to be of utmost importance to stakeholders for planning, decision making, and productive allocation of scarce resources (Goh and Law 2011; Hassani et al. 2017). Reliable forecasts also facilitate revenue allocation, direct supplier activities in the sector, and helps policymakers estimate the sector's profitability (Armstrong 1972; Álvarez-Díaz and Rosselló-Nadal 2010).

Over the years, numerous forecasting technologies and methods have been applied by practitioners and researchers alike (Song et al. (2019), but Goh and Law (2011) noted that historically, the accuracy of these methods have been mixed. Nevertheless, a more recent review of 211 key tourism demand forecasting literature published between 1968-2018 saw Song et al. (2019) concluding that the accuracy of forecasting has improved alongside the diversification of forecasting models. However, there is yet a single model that outperforms all competing models under all situations and therefore, the evolution of forecasting methods for tourism demand continues (Song et al. 2019).

Accordingly, this paper aims to propose and promote the use of new forecasting methods for tourism demand forecasting based on general non-linear "State-Dependent Models" (SDMs) (Priestley 1980b) and evaluate their performance against popular and powerful benchmark models representing both parametric and non-parametric forecasting techniques.

Very few studies have evaluated the short-term forecasting accuracy of SDMs. Cartwright (1985a) compared the forecasting performance of ARMA, Bilinear, and SDMs on the Wolfer sunspot series and IBM Daily Stock prices and found nonlinear models such as the SDM, could forecast significantly better than linear models. Cartwright (1985b) applied SDMs to estimate the missing values in the Wolfer sunspot series using the SDM approach. Cartwright and Newbold (1982) utilised SDMs to predict North Sea oil discoveries and found that non-linear forecasting methods performed significantly better than the linear models.

Bi et al. (2021) argued that tourism demand time series exhibit complicated characteristics such as nonlinearity, periodicity, and volatility, and claimed that overlooking such features equates to ignoring the dynamic temporal relationship between observations. Nonlinearity is even present in hourly tourist arrivals data (Zheng et al. 2021). Bi et al. (2021) employed time series deep learning on image processing and long short-term memory networks to forecast daily tourist arrivals of two popular Chinese tourist attractions and showed significant improvements in forecasting accuracy relative to benchmark models. Other deep learning methods can be found in Zhang et al. (2017); Radenović et al. (2018); Law et al. (2019); Bi et al. (2020). In addition, the study of structural breaks and outliers as represented by mean shifts is also important as they can significantly affect model estimation (Qiu et al. 2021). Generally, the effects of structural breaks or mean shifts can be estimated by incorporating dummy variables. Qiu et al. (2021) performed outlier smoothing to minimise the effects of outliers. In our paper, effects of outliers are not removed but we endeavour to study the overall nonlinear form in the tourism demand data through mean shifts.

In this context, the use of SDMs is advantageous as it provides an overview about specific nonlinear form. SDM is a general time series model and can be fitted without a priori knowledge of the underlying model. Also, the SDM algorithm includes, as special cases, most times series models such as Logistic Smooth Transition Autoregressive (LSTAR) and Exponential Smooth Transition Autoregressive (ESTAR) models, threshold, and nonlinear threshold autoregressive models as well as standard linear models. Also, with respect to the traditional STAR models, SDMs are more general and can handle more than two regimes.

Whilst forecasting growth in tourism demand, this paper makes several contributions to tourism demand forecasting literature. Firstly, it introduces two new and viable forecasting methods based on general non-linear "State-Dependent Models" (SDM) for tourism demand forecasting through applications to both simulated and real data. Secondly, the introduction and consideration of such new methods for tourism demand forecasting are crucial as the ongoing pandemic has shown that traditional methods might be out-of-date and ineffective (Zhang et al. 2021). Therefore, we are expanding the pool of options available for tourism demand forecasting and stimulating further research and development to improve the accuracy of tourism demand forecasts. Thirdly, we conduct extensive Monte Carlo simulation studies where the results can be of value to researchers and forecasters in terms of obtaining a better understanding of the effectiveness of the SDM techniques under varied hypothesised conditions. Fourthly, in line with most existing studies that also focus on international tourist flows (Song et al. 2019), this study provides forecasts at the disaggregated level for growth in Japanese

tourism demand based on the purpose of travel: tourism, business, and other. Fifth, the study considers various time horizons during the forecasting exercise as short-term forecasts are required for scheduling and staffing, while medium-term forecasts for planning tour operator brochures and long-term forecasts for investment in aircraft, hotels, and infrastructure (Hassani et al. 2017).

The consideration of Japanese tourist arrivals is justified as Japan is the third-largest destination in Asia (UNWTO 2019) and tourism is now recognized as a key contributor to the Japanese economy with the total contribution of travel and tourism to GDP being 7% of the total economy in 2019 whilst the total contribution of the sector to employment had risen to 8% of total employment (World Travel & Tourism Council 2020).

In terms of the benchmark models, these have been widely applied in the context of tourism demand forecasting both historically and recently (see, for example, (Law 2000; Chen et al. 2009; Athanasopoulos and de Silva 2012; Bergmeir et al. 2016; Cho 2016; Witt et al. 2016; Silva et al. 2019).

The remainder of this paper is organized such that Section 2 describes the methodology and estimation of the State-Dependent Models. Section 3 presents the results from Monte Carlo simulation studies. Section 4 presents summary statistics for the tourist arrival data in Japan alongside the metrics used for forecast evaluation. Section 5 analyses the forecasting results with concluding remarks and suggestions for future research in Section 6.

2. State-Dependent Models (SDMs)

2.1 Estimation Procedure

Priestley (1980b) developed a general class of non-linear models which includes the linear and many specific non-linear time series models. In this section, the parameter estimation of SDMs is briefly described, before the three forecasting procedures with SDM are presented. SDM is a recursive and efficient approach and it works well especially in the case of data with no prior assumptions about the type of nonlinearity (Priestley 1980b) in the data. This general non-linear model contains the linear time series as well as many types of non-linear models (such as the Exponential Autoregressive Models, Threshold Autoregressive Models, Nonlinear threshold autoregressive models, Logistic and Exponential Smooth Transition Autoregressive Models). For more extensive theoretical discussions and applications of this model, see Priestley (1980b) and Haggan et al. (1984).

Consider the following linear AR(*k)* model:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_k X_{t-k} + \epsilon_t \qquad (2.1)$$

All the coefficients $\{\boldsymbol{\phi_u}: u = 0, 1, \ldots, k\}$ are constants.

The SDM is an extension of the linear AR time series model in which the coefficients in (2.1) become functions of the relevant past information. This extension of the linear AR model leads to the general non-linear model as:

$$X_t = \phi_0(\boldsymbol{x_{t-1}}) + \phi_1(\boldsymbol{x_{t-1}})X_{t-1} + \cdots + \phi_k(\boldsymbol{x_{t-1}})X_{t-k} + \epsilon_t. \qquad (2.2)$$

In (2.2), the vector $\boldsymbol{x_{t-1}}$ contains the relevant past information and is given by:

$$\boldsymbol{x_{t-1}} = (1, X_{t-k}, \ldots, X_{t-1})$$

The following shows how different types of models are encompassed by the SDM using particular forms of $\phi_0, \phi_1, \phi_2, \ldots, \phi_k$.

(a) Linear Models

Take $\{\phi_0, \phi_1, \phi_2, \ldots, \phi_k\}$ to be constant and independent of $\boldsymbol{x_{t-1}}$.

(b) Exponential AR models

Take $\phi_0 = 0$, and for $u = 1, 2, \ldots, k$,

$$\phi_u(\boldsymbol{x_{t-1}}) = \theta_u + \pi_u e^{-\gamma_i X_{t-1}^2}$$

(c) Threshold AR models

For $u = 0, 1, 2, \ldots, k$,

$$\phi_u(\boldsymbol{x_{t-1}}) = \begin{cases} \theta_u^{(1)}, & if\, X_{t-d} \le c \\ \theta_u^{(2)}, & if\, X_{t-d} > c \end{cases}$$

(d) Non-linear Threshold AR

Take $\phi_0 = 0$, and for $u = 1, 2, \ldots, k$,

$$\phi_u(\boldsymbol{x_{t-1}}) = \begin{cases} \theta_u + \pi_u|X_{t-d}|, & if\, X_{t-d} \le c \\ \theta_u + \pi_u c, & if\, X_{t-d} > c \end{cases}$$

The Non-linear threshold AR model was proposed by Ozaki (1981)

(e) Logistic Smooth Transition models (LSTAR)

For $u = 0, 1, 2, \ldots, k$,

$$\phi_u(\boldsymbol{x_{t-1}}) = \theta_u + (\pi_u - \theta_u)\left[1 + e^{-\gamma_i(X_{t-1}-c)}\right]^{-1}$$

(f) Exponential Smooth Transition models (ESTAR),

For $u = 0, 1, 2, \ldots, k$,

$$\phi_u(\mathbf{x_{t-1}}) = \theta_u + (\pi_u - \theta_u)[1 - e^{-\gamma_i(X_{t-1}-c)^2}]$$

To fit a SDM, the parameters ϕ_0, ϕ_1, …, ϕ_k need to be estimated recursively and depend on the state vector $\boldsymbol{x_{t-1}}$. Assuming the coefficients in (2.2) are locally represented as a smooth and linear function of the vector $\boldsymbol{x_t}$, the updated equations for the parameters can be expressed as:

$$\phi_u(\boldsymbol{x_t}) = \phi_u^{(0)} + \boldsymbol{x_t \beta_u}, \qquad (2.3)$$

where $\boldsymbol{\beta_u}$ are gradient parameters and assume to follow a random walk model. Alternatively, equation (2.3) can be written as:

$$\phi_u(\boldsymbol{x_{t+1}}) = \phi_u(\boldsymbol{x_t}) + \boldsymbol{\Delta x_{t+1} \beta_u^{(t+1)}},$$

where $\boldsymbol{\Delta x_{t+1}} = \boldsymbol{x_{t+1}} - \boldsymbol{x_t}$.

Priestley (1980) gave a reformulation of SDM in a state-space form in which the state-vector $\boldsymbol{\theta_t}$ includes all the current parameters. Applying the extended Kalman algorithm to the reformulated formula, the coefficients can be estimated recursively as follows:

$$\widehat{\boldsymbol{\theta}}_{\boldsymbol{t}} = \boldsymbol{F_{t-1}}\widehat{\boldsymbol{\theta}}_{\boldsymbol{t-1}} + \boldsymbol{K_t}\{X_t - \boldsymbol{H_t F_{t-1}}\widehat{\boldsymbol{\theta}}_{\boldsymbol{t-1}}\}, \qquad (2.4)$$

where $\boldsymbol{H_t} = (1, X_{t-1}, \cdots, X_{t-k}; 0, 0, \cdots, 0)$

$$\boldsymbol{\theta_t} = \left(\phi_0^{(t-1)}, \phi_1^{(t-1)}, \cdots, \phi_k^{(t-1)}; \boldsymbol{\beta_0^{(t)'}}, \boldsymbol{\beta_1^{(t)'}}, \cdots, \boldsymbol{\beta_k^{(t)'}}\right)'$$

and the transition matrix is defined as follows:

$$\boldsymbol{F_{t-1}} = \begin{pmatrix} & \boldsymbol{\Delta x_{t-1}} & 0 & 0 \\ I_{k+1} & 0 & \ddots & 0 \\ & 0 & 0 & \boldsymbol{\Delta x_{t-1}} \\ 0 & & \mathrm{I}_{k(k+1)} & \end{pmatrix}$$

where, $\boldsymbol{\Delta x_{t-1}} = (X_{t-1} - X_{t-2}, X_{t-2} - X_{t-3}, \ldots\ldots, X_{t-k} - X_{t-k-1})$ and $\boldsymbol{K_t}$ is called the "Kalman Gain". The observation equation (2.2) can also be written as:

$$X_t = \boldsymbol{H_t}\, \widehat{\boldsymbol{\theta}}_{\boldsymbol{t}}. \qquad (2.5)$$

To start the recursion equation (2.4), initial estimates are needed for the parameters, the variance-covariance matrix of the parameters and the residual variance, σ_ϵ^2. These can be obtained by using the first stretch of the data and fitting the linear model (2.1). The initial "gradient" parameters can be also set to 0, thus:

$$\hat{\theta}_{\mathrm{t}_0-1} = \left(\hat{\phi}_0, \hat{\phi}_1, \cdots, \hat{\phi}_k, 0, \cdots, 0\right)'.$$

Having the initial estimates of the parameters, the variance-covariance matrix, and σ_ϵ^2, the successive values of the parameters can be obtained employing the standard recursive equations for the Kalman Filter.

SDM fitting requires the setting of a parameter called the "smoothing parameter". Based on experimenting with different values of the smoothing parameters, this should ideally be between 0.0001 to 0.01. Too large values of the smoothing parameter will make estimated values of coefficients explode, and too small a value will make it difficult to detect nonlinearity (Haggan et al. 1984). However, so long as the value of the smoothing parameter is selected between 0.0001 and 0.01, as recommended by Haggan et al. (1984), the results do not differ significantly.

2.2 Forecasting procedures with SDM

Here, we propose three different SDM forecasting methods, which are outlined below.

a) The classical recursive forecasting method

Like other forecasting methods, the forecasting performance of SDM depends on fitting the appropriate SDM scheme. Given observations on $\{X_{(t)}\}$ up to the end of series (time T) and using the SDM model with the state vector as $\boldsymbol{x_{T-1}} = (1, X_{T-1}, \dots, X_{T-k})$, one step ahead forecasts, $\hat{X}_{T+1}$ can be computed from (2.5) by:

$$\hat{X}_{T+1} = \hat{\phi}_0^{(t)} + \hat{\phi}_1^{(t)} X_T + \cdots + \hat{\phi}_k^{(t)} X_{T-k+1} = \boldsymbol{H_{T+1}\widehat{\theta}_{T+1}} \qquad (2.6)$$

where $H_{T+1} = (1, X_T, \dots, X_{T-k+1}, 0, 0, \dots, 0)$, and $\boldsymbol{\widehat{\theta}_{T+1}}$ is computed recursively as

$$\boldsymbol{\widehat{\theta}_{T+1} = F_T \widehat{\theta}_T},$$

where $$\boldsymbol{F_T} = \begin{pmatrix} & \boldsymbol{\Delta x_T} & 0 & 0 \\ I_{k+1} & 0 & \ddots & 0 \\ & 0 & 0 & \boldsymbol{\Delta x_T} \\ 0 & & \mathrm{I}_{k(k+1)} & \end{pmatrix}.$$

Parameters $\hat{\phi}_0, \hat{\phi}_1 \dots, \hat{\phi}_k$ in (2.6) and the gradients $\boldsymbol{\gamma_0^{(t)}, \gamma_1^{(t)}, \dots, \gamma_k^{(t)}}$ are updated as follows:

$$\phi_u^{(T)} = \phi_u^{(T-1)} + \boldsymbol{\Delta x_T \gamma^{(T)}}$$
$$\boldsymbol{\gamma^{(T+1)} = \gamma^{(T)}}$$

Similarly, two-steps ahead forecasts can be obtained from:

$$\hat{X}_{T+2} = \boldsymbol{H_{T+2}.\widehat{\theta}_{T+2}},$$

where $\boldsymbol{\widehat{\theta}_{T+2} = F_{T+1}\widehat{\theta}_{T+1}}$, $H_{T+2} = \{1, \hat{X}_{T+1}, X_T, \dots, X_{T-k+2}, 0, 0, \dots, 0\}$

and the gradient parameters remain the same as the last available estimated gradients, i.e.:

$$\boldsymbol{\gamma^{(T+2)} = \gamma^{(T+1)} = \gamma^{(T)}}.$$

Similarly, the *h*-steps ahead forecasts can also be computed recursively as:

$$\boldsymbol{\widehat{X}_{T+h} = H_{T+h}.\widehat{\theta}_{T+h}} \qquad (2.7)$$

with the gradient parameters after the time T (end of the series) again updated as:

$$\boldsymbol{\gamma}^{(\boldsymbol{T+h})} = \boldsymbol{\gamma}^{(\boldsymbol{T})}$$

As can be seen above, the updating procedure for the gradient parameters after *T* is different from the updating procedure before *T,* which is to allow the gradients $\boldsymbol{\gamma}_{\boldsymbol{u}}^{(\boldsymbol{t})}$ to wander in the form of "random walk", and update them, for each *t*, to minimise the discrepancy between the observed value X_{t+1} and its predicted value, $\hat{X}_{t+1}$.

b) Estimating the parameters from unsmoothed surface, evaluated with the most recent forecasts

As shown in the previous section, between time *T* and $T + h$, the coefficients of $\hat{\phi}_0$, $\hat{\phi}_1 \ldots$, $\hat{\phi}_k$ are strictly linear functions of the state vector, as the gradients remain unchanged. However, for long-term forecasting, assuming strict linearity in the parameters is not desirable (Priestley 1980a). Therefore, in this case, it may be better to estimate the parameters from the surfaces fitted up to time *T*, but evaluated with the most recent forecast values (Priestley 1980a). Thus, if the parameter surfaces fitted up to time *T* are denoted by $\hat{\phi}_u^{(T)}(\boldsymbol{x_{T-1}})$, then the *h*-steps ahead forecast can be similarly obtained from (2.7) with the parameter estimated as:

$$\hat{\phi}_u = \phi_u^{(T)}(\hat{\boldsymbol{x}}_{\boldsymbol{T-1+j}}), \qquad j = 1, 2, \ldots, h$$

Using this method, the forecasts are no longer based on parameters constrained to be linear functions of the state vector**.**

c) Estimating the parameters from the smoothed surfaces

Computing the forecasts for this case is like the method described in the previous section. However, $\hat{\phi}_u^{(t)}$ are computed from the smoothed surface of the parameters. The motivation behind applying this method is to estimate the parameters from a smoothed surface to obtain more accurate values for the parameters through a robust fitting procedure which reduces the influence of extreme values on the parameters. Through including the locally weighted polynomial regression and smoothing scatter plots (LOWESS) method, the estimates of the parameters were obtained from the smoothed surface, using a grid search technique. The $(f*n)^{\text{th}}$ nearest neighbours are used and weighted least squares estimation is implemented in

calculating the smoothed values. The weights are proportional to the distance between the scatterplot points and the fitted values. There are four parameters for the locally weighted regression: the order of polynomial d, the function W used to determine weights, the number of iterations t for robust model fitting and the fraction of data f used to determine the amount of smoothing. Cleveland (1979) recommended using $d = 1$ ensuring adequate smoothing and good computational ease. In terms of the function, the tricube function was argued to provide adequate smoothing. Two iterations are good for almost all situations, and he suggested that in general choosing $f = 0.2$ to 0.8 is suitable. In this study, tricubic functions are used for polynomial fitting, t and d were set as the default values of 3 and 1 .

3. Monte Carlo Simulation Study

The purpose of this Monte Carlo Simulation study is to apply the state dependent estimation approach on data generated from linear and various non-linear models, and subsequently to demonstrate the effectiveness of the SDM estimation method. This section considers several different simulations, including a linear $AR(2)$ process, logistic smooth transition autoregressive (LSTAR) and exponential smooth transition autoregressive (ESTAR) processes. In the linear AR process, the coefficients are constant over time, and for LSTAR and ESTAR non-linear models, the coefficients are logistic and exponential functions of the transition variable, respectively. The SDM algorithm was then applied on the simulated data from different models. In the following section, the solid lines in each graph represent the functional form of the parameter for the true (non)linear model. The dashed lines represent the parameter estimated from SDM. The order of the models is chosen by minimising the BIC criterion of fitting a linear model to data. To obtain the starting values for the recursion, a linear AR model was fitted to the first *m* observations and the recursion began at *m/2*, as recommended by Priestley (1980b). The effectiveness of the SDM is examined by graphical comparison of the true functional form of the parameter (solid line) with the estimated parameter computed by the SDM algorithm and smoothed by a non-parametric function fitting technique (dashed line). It is important to note that the SDM algorithm will not necessarily reproduce the "natural" form of the parameters. This is because the SDM approach is based on a local linearization and therefore the coefficients in the general model of the form given below

$$X_t = h(X_{t-1}, X_{t-2}, \dots, X_{t-k}) + \epsilon_t$$

will follow the form of the first partial derivatives of *h* with respect to X_{t-u} ,

i.e. $\hat{\phi}_u = \frac{\delta \mathrm{h}}{\delta X_{t-u}},\ u = 1, 2, \dots, k,$ (Priestley 1980a).

3.1 Linear Model Simulation

In this simulation study, 500 observations from the following linear $AR(2)$ process were generated:

$$X_t = 0.4X_{t-1} - 0.7X_{t-2} + \epsilon_t,$$

where $\{\epsilon_t\}$ is a sequence of independently and normally distributed random errors with a mean of zero and a standard deviation of one. In this case, the first stretch of the data, (first 50 observations) were used and a linear $AR(2)$ model was fitted to produce the initial values of $\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2$, the variance-covariance matrix, and variance of the residuals, σ_ϵ^2. The SDM algorithm was then applied, and the recursion started from $t = 25$ and then updated over time applying the extended-Kalman Filtering. Writing X_t in the form $X_t = h(X_{t-1}, X_{t-2}) + \epsilon_t$, and partially differentiating h with respect to X_{t-1} and X_{t-2}, the functional form of ϕ_1 and ϕ_2 are given as: $\hat{\phi}_1 = 0.40$ and $\hat{\phi}_2$ = -0.70. Thus, in the case of linear models, $\hat{\phi}_u$ will follow the natural form.

Figure 1 shows the resulting graphs of the parameters of $\hat{\phi}_0$, $\hat{\phi}_1$ and $\hat{\phi}_2$ estimated from SDM and plotted against the transition variable, X_{t-1}.

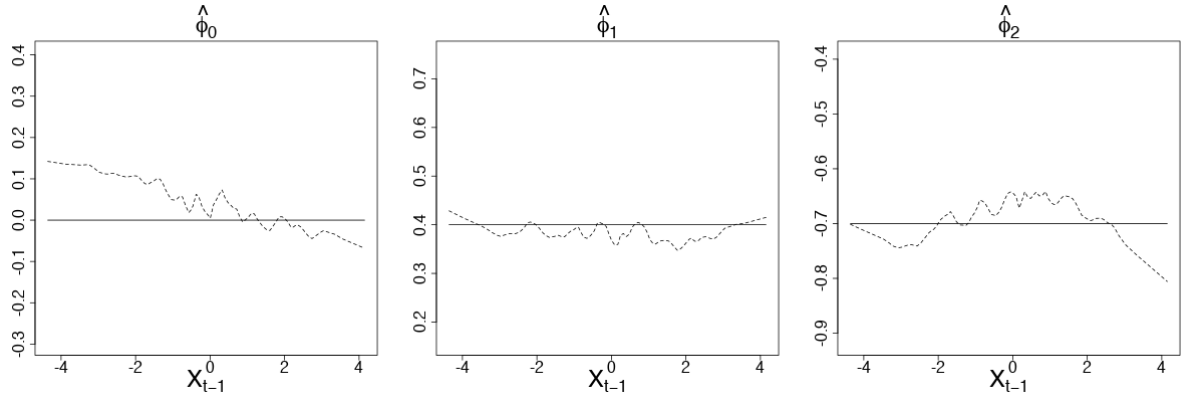


Figure 1: Graphs of the estimated and true parameters for the linear model.

From a visual inspection of the parameters $\hat{\phi}_0$, $\hat{\phi}_1$ and $\hat{\phi}_2$, it was observed that the SDM algorithm has reproduced the true shape of the parameters, with $\hat{\phi}_0$ staying close to 0.00; $\hat{\phi}_1$was very close to 0.40, and $\hat{\phi}_2$ fluctuated around -0.70. Therefore, based on the data only, the SDM model provided strong evidence that the series was generated by a linear model.

3.2 Logistic Smooth Transition Autoregressive (LSTAR) Simulation

In this case, again, 500 observations were generated from the following LSTAR model. The first 50 observations were used for initialisation, and the recursion of the SDM started from $t = 25$.

$$X_t = 0.4(1 + e^{-X_{t-1}})^{-1} X_{t-1} + \epsilon_t.$$

Writing X_t in the form $X_t = h(X_{t-1}) + \epsilon_t$, and partial differentiating h with respect to X_{t-1}, the functional form of ϕ_1 is given as:

$$\phi_1(X_{t-1}) = \frac{2(1 + e^{-X_{t-1}} + X_{t-1}e^{-X_{t-1}})}{5(e^{X_{t-1}} + 1)^2}.$$

This function was plotted on a solid line in Figure 2, to represent the true functional form of ϕ_1. The following figure shows graph of $\hat{\phi}_1$ estimated using SDM plotted against X_{t-1}.

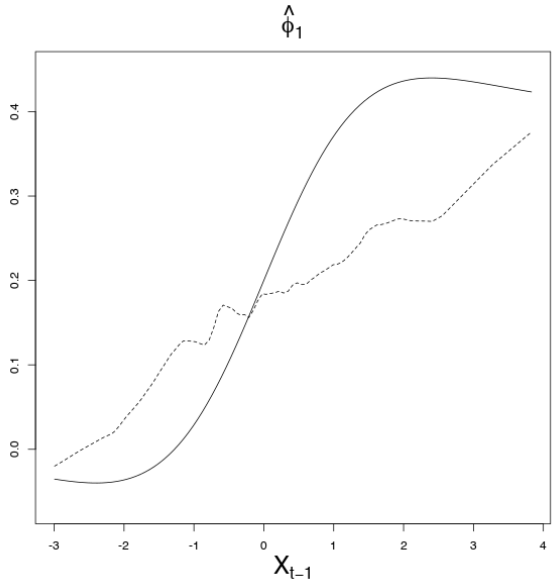


Figure 2: Graph of $\hat{\phi}_1$ against X_{t-1}

In Figure 2, the estimated line (in dashed form) produced by the SDM algorithm pointed us clearly in the direction of the LSTAR model. The range of fitted values of $\hat{\phi}_1$ were approximately the same as the true parameter, which ranged from 0.00 to 0.40 and shows an upward trend. The slope of the estimated function was slightly less than the true parameter. The intercept of the curve is also very close to the true parameter. The SDM algorithm estimated

the linear part of the model very accurately and the estimated value of $\hat{\phi}_1$ is close to the true value of the parameter (0.20), when X_{t-1} is equal to zero. We also generate data form LSTAR models with a downward slope, the SDM fitting results (available from the authors upon request) were consistent, showing a downward sloping trend in the plot when $\hat{\phi}_1$ parameter is plotted against the state dependent variable, X_{t-1}.

Distinguishing between ESTAR and LSTAR models are important, therefore in the next section we generate data with ESTAR characteristics and investigate whether the SDM can reproduce the correct nonlinear symmetric shape of the coefficient when is plotted against the X_{t-1}, and to see whether it is symmetric around zero.

3.3 Exponential Smooth Transition Autoregressive (ESTAR) Simulation

In this example, 500 observations were generated using the model:

$$X_t = 0.7\left(1 - e^{-1.1X_{t-1}^2}\right)X_{t-1} + \epsilon_t.$$

The SDM algorithm was run to produce a plot of estimated $\hat{\phi}_0$, $\hat{\phi}_1$ against X_{t-1}. Again, the first 50 observations were used to produce initial values of the parameters μ, ϕ, the variance-covariance matrix, and σ_ϵ^2 through fitting a linear autoregressive model, and recursion started from $t = 25$. Figure 3 shows the graph of $\hat{\phi}_1$ estimated via SDM and plotted against X_{t-1}.

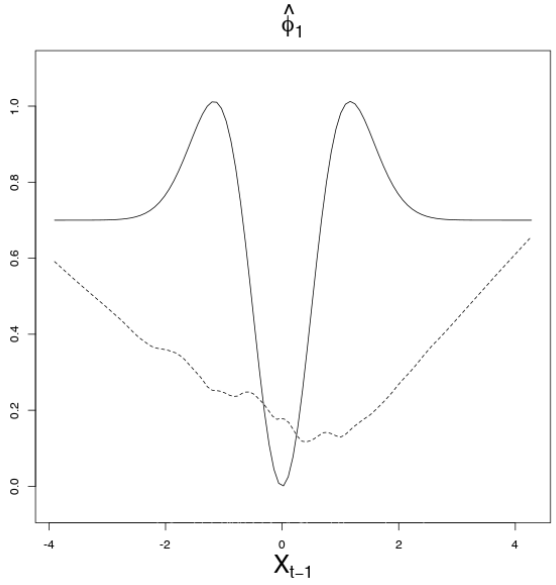


Figure 3: Graph of $\hat{\phi}_1$ against X_{t-1}

The actual ϕ values should vary according to the first derivative of the ESTAR model, i.e.,

$$\hat{\phi}_1(X_{t-1}) = \frac{e^{-1.1X_{t-1}^2}(35e^{1.1X_{t-1}^2} + 77X_{t-1}^2 - 35)}{50}.$$

Figure 3 shows that the fitted curve has very similar characteristics with the true value of ϕ_1, the estimated parameter is approximately correct, and it has similar shape to the true curve.

The values of $\hat{\phi}_1$ for large values of X_{t-1} at the two sides of the graph are very close to the expected value, 0.70, and the minimum of the fitted curve also occurs near the point where $X_{t-1} = 0.00$, i.e., the linear part of the model was also estimated accurately. Therefore, the SDM was able to provide a very good fit through prescribing the right non-linear shape and the right range of values. In particular, the symmetrical appearance of Figure 3 is obviously different from other well-known models, e.g., Linear, LSTAR, Threshold Autoregressive (TAR), and Bilinear models, and clearly pointed us in the direction of ESTAR models. Therefore, without any a priori knowledge of the underlying model, again the SDM fitting algorithm captured the non-linearity in the data, estimated the model correctly and distinguished the difference between ESTAR and other types of models.

3.4 Simulating Model with a Mean Shift (Change in Intercept)

As tourism demand in Japan shows a shift in the level, we examine whether the SDM algorithm can capture the mean shifts in the data using a Monte Carlo simulation study. 500 observations were generated again, and the first 50 observations were used to obtain the initial values for the parameters of $\hat{\phi}_0$, $\hat{\phi}_1$. The SDM algorithm was applied, and recursion started from the middle of the data at $t = 25$ using the initial values of the parameters. The figure (Figure 4) presented below shows the time series plot for the data generated from the following model:

$$X_t = (1 + e^{-\gamma z})^{-1} + 0.6X_{t-1} + \epsilon_t, \text{ with } \gamma = 10.0,$$

where, γ represents the rate of adjustment in the logistic function, z takes its value from -5 to 5 at equal intervals of 0.02 (for 500 observations). It can be observed from Figure 4 that the series demonstrates an abrupt change in the level (with γ = 10.0) at time t = 250 (at the middle of data).

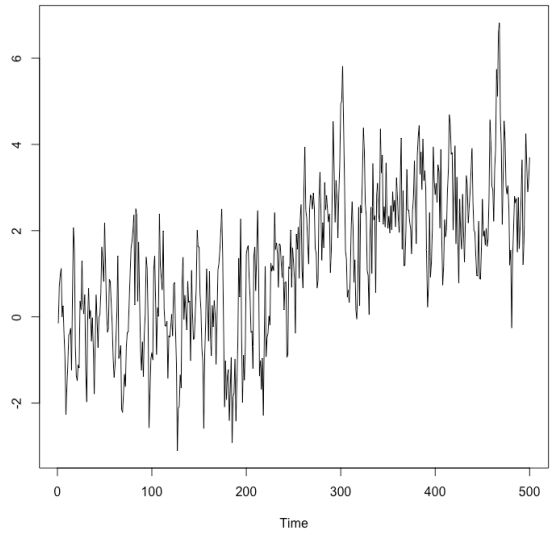


Figure 4: Graph of the simulated time series for γ = 10.0

Figure 5 contains graphs of the estimated values for the parameters $\hat{\phi}_0$, and $\hat{\phi}_1$, plotted against time and X_{t-1}. The solid lines represent the true values simulated. The preliminary estimation gave the following initial values for the parameters:

$\hat{\phi}_0$ = -0.00580, $\hat{\phi}_1$ = 0.615.

Figure 5 shows a fast adjustment in the value of $\hat{\phi}_0$ parameter at time *t* =250. This indicates that the SDM algorithm correctly estimated the jump in the data, with a clear shift in the parameter from zero to one. The estimated values are very close to the true functional form. The second graph in Figure 5 demonstrates that estimate of ϕ vary around the simulated value of 0.6. This shows that reliable estimates of parameters $\hat{\phi}_0$, and $\hat{\phi}_1$ can be obtained using the SDM technique. The results provided strong evidence that the data were generated from a linear model with a shift in the mean.

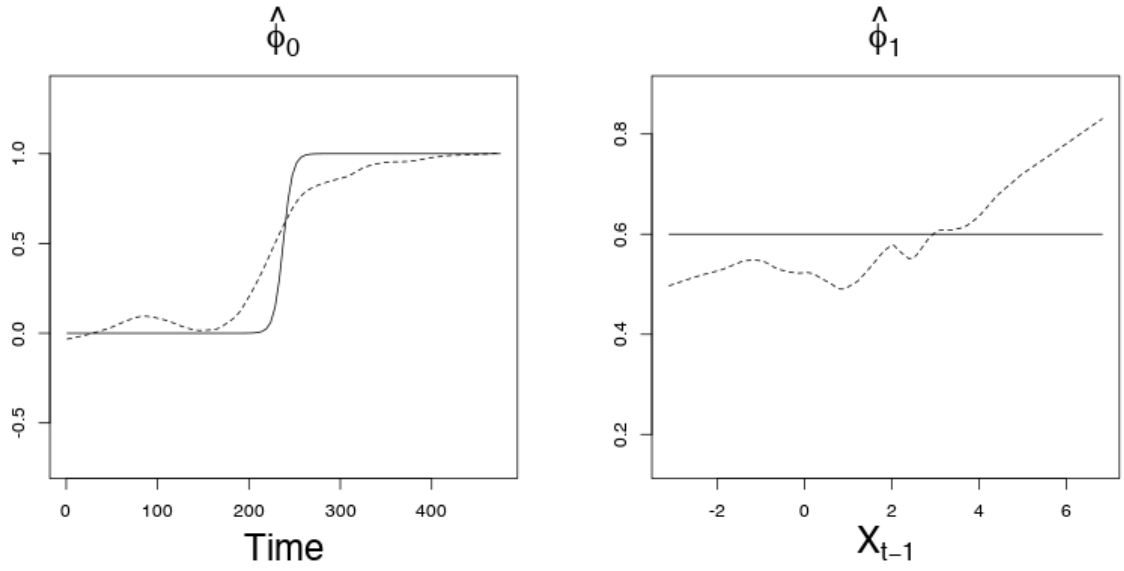


Figure 5: Plot of $\widehat{\boldsymbol{\phi}}_{\mathbf{0}}$ against time and $\widehat{\boldsymbol{\phi}}_{\mathbf{1}}$ against $\boldsymbol{X_{t-1}}$ with $\boldsymbol{\gamma}$ = 10.0.

3.5 Simulation of a Model with Mean Shift (Change in Intercept) & Change in Slope

We further examined the SDM estimation procedure, simulating 500 observations from a time series model that includes both shift in the mean and non-linearity in the autoregressive parameter. Figure 6 shows the data generated form the following model with a smooth shift in the mean.

$$X_t = -(1 + e^{-0.50\,z})^{-1} + 0.5(1 + \mathrm{e}^{-1.0X_{t-1}})^{-1}X_{t-1} + \epsilon_t$$

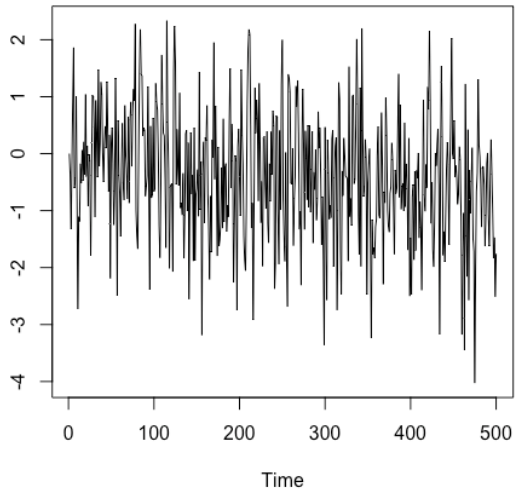


Figure 6: Graph of the simulated time series for $\boldsymbol{\gamma_1}$ = 0.50 & $\boldsymbol{\gamma_2}$ = 1.00.

Like other cases, again we used the first 50 observations to obtain the initial estimate of the parameters, and recursions started from t = 25.

$\hat{\phi}_0$ = -0.05528, $\hat{\phi}_1$ = 0.13958.

Figure 7 presents the estimated parameter $\hat{\phi}_0$ plotted against time, showing a smooth downward trend. The second graph in Figure 7 shows $\hat{\phi}_1$ plotted against X_{t-1}. The estimated values in both graphs are close to the true functional forms. The results show that the SDM algorithm can indeed capture the changes in both parameters and indicated that the data were generated from a non-linear (LSTAR) model with a smooth shift in the mean.

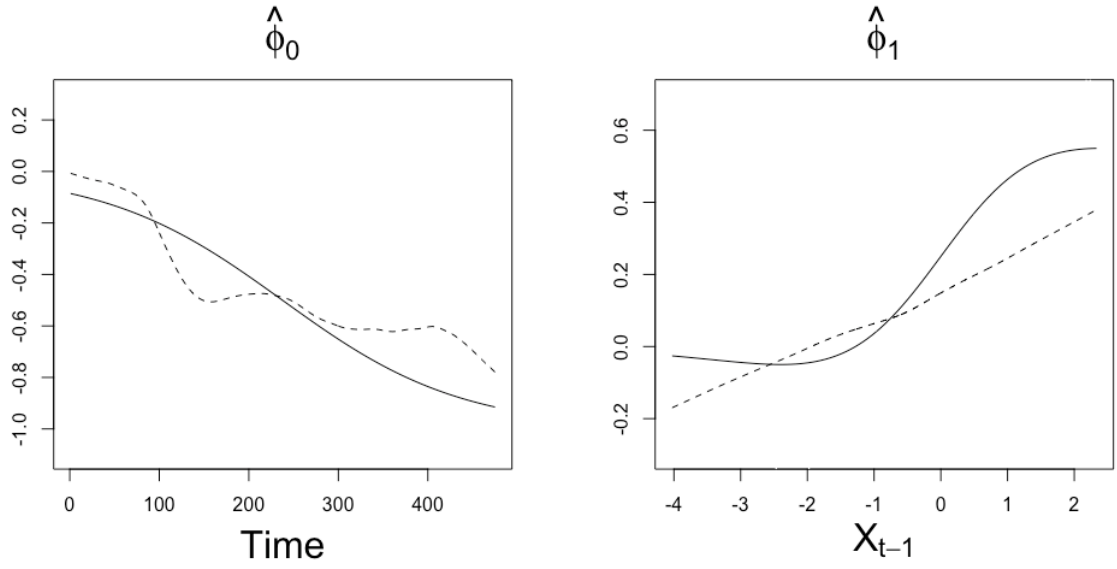


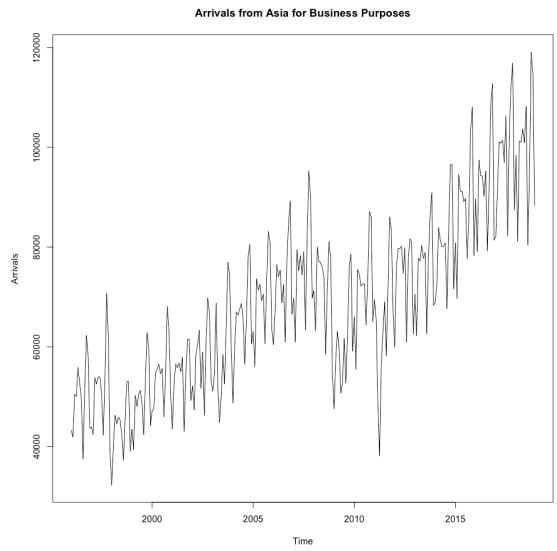
Figure 7: Plot of $\widehat{\boldsymbol{\phi}}_{\mathbf{0}}$ against time and $\widehat{\boldsymbol{\phi}}_{\mathbf{1}}$ against $\boldsymbol{X_{t-1}}$ with $\boldsymbol{\gamma_1}$ = 0.50 and $\boldsymbol{\gamma_2}$ = 1.00.

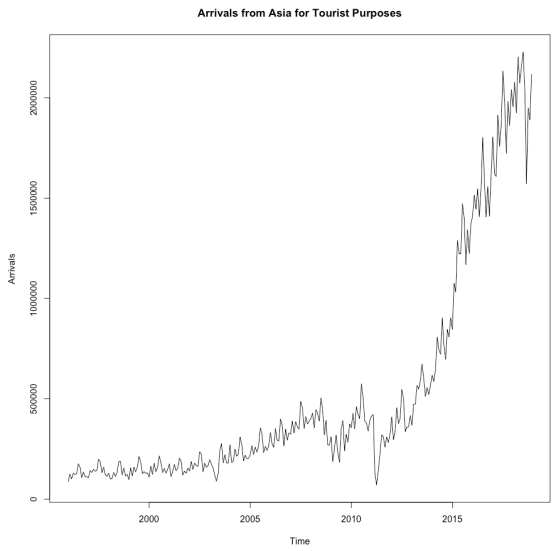
4. The Data and Forecast Accuracy Metrics

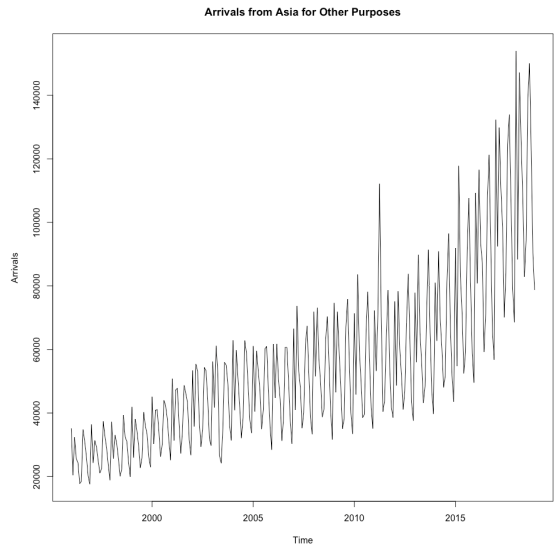
4.1 The Data

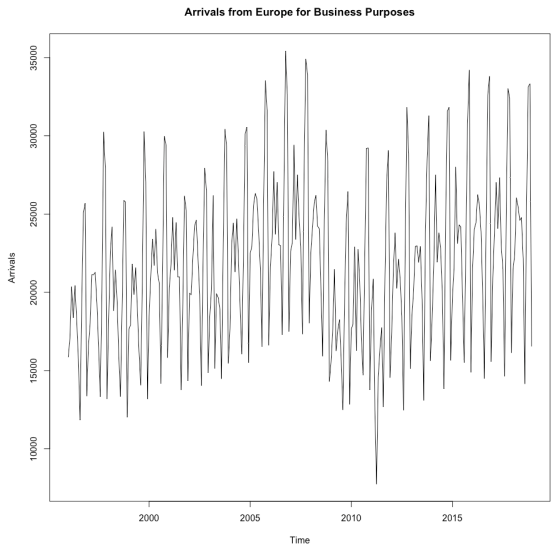
The data in the study were taken from the Japan Tourist Board Research & Consulting Co. website[1]. This study attempts to forecast the "Grand total" as well as continent-specific time series - Asian, European, and North American disaggregated by purpose of travel, "Tourist", "Business" and "Other". The data in this research are monthly arrivals ranging from January 1996 to December 2018. Figure 8 presents the time series plots of the inbound tourist arrivals for tourism, business, and other purposes for Asia, Europe, North America, and the Grand Total. Almost all tourist arrivals for tourism purposes showed an increasing trend, with a decline during the banking crisis. It is noteworthy that even though the time span of this data does not capture the effects of the COVID-19 pandemic, it remains influenced by the effects of the external shock of the banking crisis which had a similar structural impact on the series. Of course, the impact would be far worse for tourism during the pandemic as borders were closed indefinitely. As the pandemic is ongoing, thereby reducing access to data that can be credibly compared as pre- and post-pandemic, it makes sense to rely on the current timelines. Dominant seasonality was also evident in all series, except tourism from Asia.

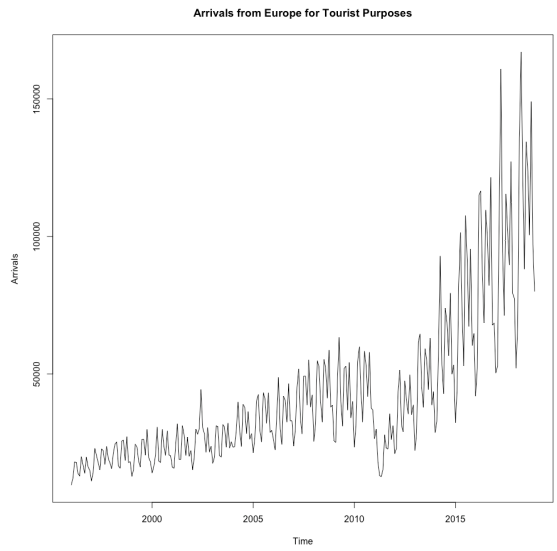
[1] Data collected from JTB Tourism Research & Co website - https://www.tourism.jp/en/tourism-database/stats/

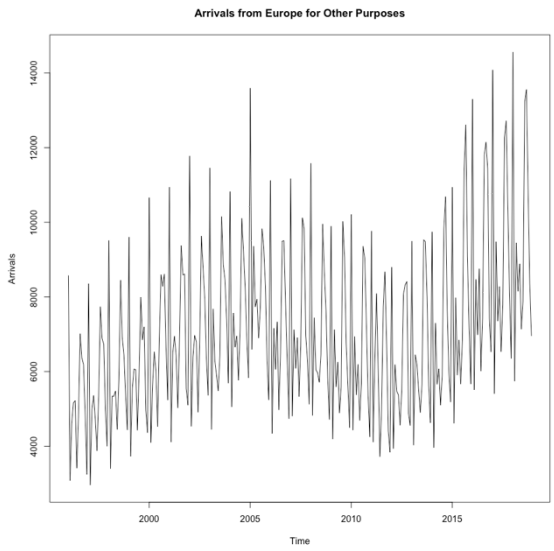

Arrivals from Asia for Business Purposes
Arrivals
Time


Arrivals from Asia for Tourist Purposes
Arrivals
Time


Arrivals from Asia for Other Purposes
Arrivals
Time


Arrivals from Europe for Business Purposes
Arrivals
Time


Arrivals from Europe for Tourist Purposes
Arrivals
Time


Arrivals from Europe for Other Purposes
Arrivals
Time

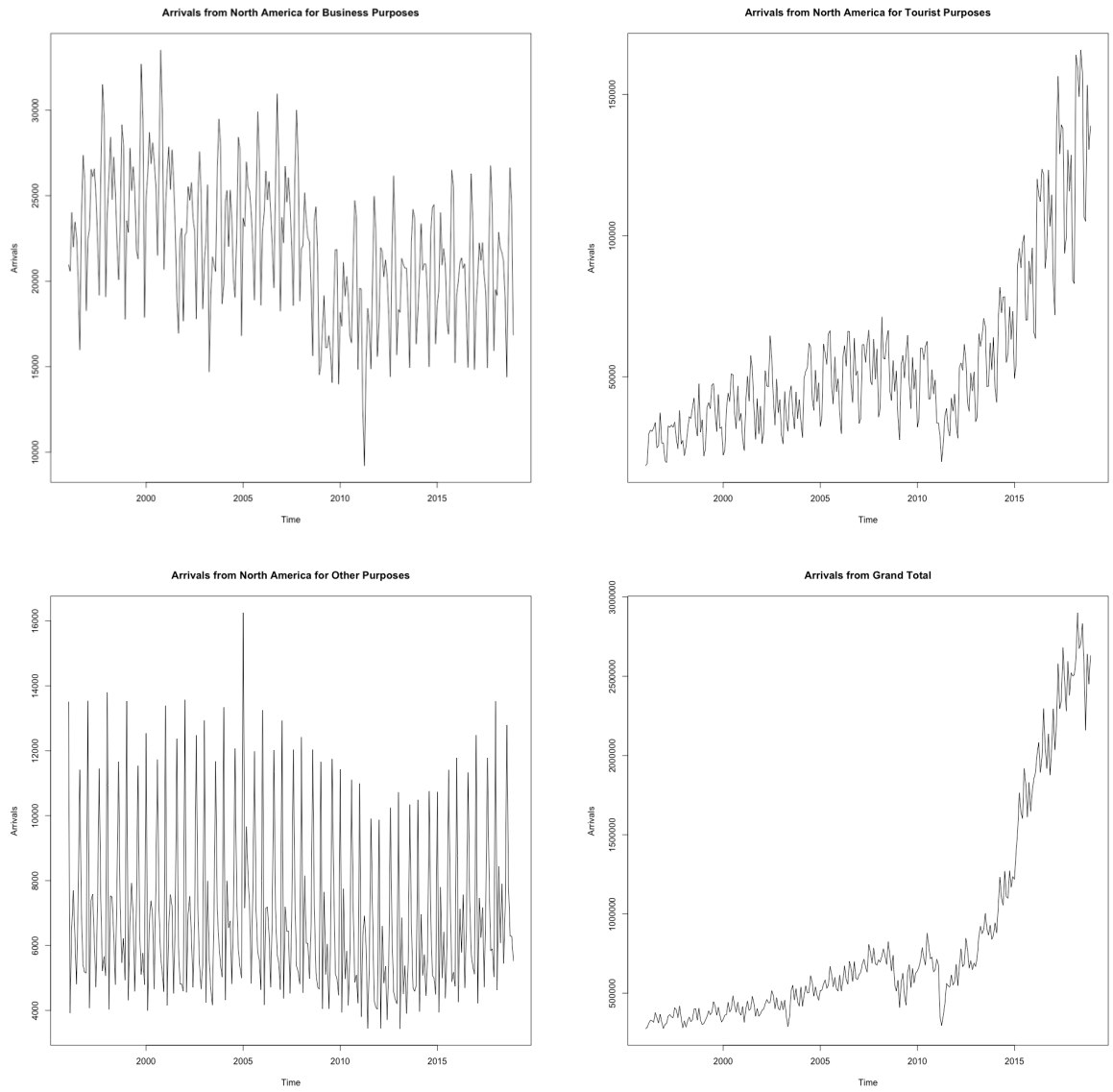


Figure 8: Graphs of arrivals for different purposes

To remove the seasonality, we transformed the data into $\log(X_t) - \log(X_{t-12})$ i.e., annual percentage change was employed in the forecasting section for all the models. This is an approach that has been adopted in tourism demand forecasting literature in the past. For example, Sheldon (1993) studied the percentage changes in international tourist expenditure and arrivals for 15 OECD countries whilst Goh and Law (2002) also relied on percentage change in tourist arrivals within their forecasting study. Finally, Goh et al. (2008) also transformed the continuous values of time series into percentage changes within their work.

Table 1 provides summary statistics of the percentage of annual change for the ten arrivals time series data. Seasonal R^2 was also computed to check the extent of the seasonal variations in the data. The seasonal R^2 in the table was computed by regressing the monthly

change in arrivals against twelve seasonal dummies. The Seasonal R^2 for all series, range from 47.9% to as high as 94.6%. In terms of annual growth rates, travel for the purposes of tourism increased far more than business and other purposes, with the highest growth experienced for tourists from Asia followed by Europe and North America. Table 1 shows that the number of Business passengers travelling from North America has declined during this period. Overall, the grand total shows an average growth of around 9.5% per year during this period. As can be seen from the graphs, the growth is particularly evident and sharp from 2012, few years after the banking crisis of 2008. The volatility of the 'European Others' series is the highest amongst all regions, followed by Asian tourism and the Grand Total.

Table 1: Summary statistics of annual growth of tourist arrivals.

Category	Mean	Standard deviation	Seasonal R^2	Coefficient of Variation
Asia Tourist	12.72	32.69	0.4791	2.569
Asia Business	3.158	14.07	0.7317	4.457
Asia Others	6.893	10.40	0.9237	1.509
North America Tourist	7.054	19.55	0.8546	2.772
North America Business	-0.277	14.30	0.8114	51.666
North America Others	0.439	10.31	0.9601	23.454
Europe Tourist	8.760	26.56	0.8108	3.032
Europe Business	1.201	14.92	0.8964	12.428
Europe Others	0.497	39.460	0.9463	79.345
Grand Total	9.534	20.857	0.577	2.188

4.2 Metrics

We rely on the out-of-sample Root Mean Square Errors (RMSE) and Relative Root Mean Square Errors (RRMSE) to measure forecasting performance. These metrics have been used in tourism forecasting literature both recently and historically, see for example Silva et al. (2019), Hassani et al. (2015) and references therein. The RMSE for *k*-steps ahead forecasts can be defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum\left(\hat{X}_{t+k} - X_{t+k}\right)^2},$$

where $\hat{X}_{t+k}$ is the *k*-step ahead forecast computed by the ARIMA, ETS, NN, LSTAR, ESTAR, TVP models, and the three versions of the SDM models, *N* is the number of observations, and

X_{t+k} is the actual value at a given horizon. Likewise, the formula for the RRMSE can be defined as follows where we consider the Naïve model as the benchmark:

$$Relative\ RMSE = \frac{RMSE\ from\ a\ competing\ model}{RMSE\ from\ Naïve\ model}.$$

Thus, a RRMSE value of less than 1.0 indicates that the competing model can forecast better than the benchmark by 1-RRMSE%.

The accuracy of competing forecasts are evaluated for statistically significant differences using the Kolmogorov-Smirnov Predictive Accuracy (KSPA) test in Hassani and Silva (2015). In brief, the KSPA test exploits the two-sample two-sided Kolmogorov-Smirnov test to determine the existence of a statistically significant difference between the distributions of two forecast errors (Hassani and Silva 2015). The null hypothesis is that two forecast errors share the same distribution. This test has been used recently in various forecasting literature including tourism forecasting, see for example, He et al. (2021), Tian et al. (2021), Huang and Hao (2021), Fan et al. (2021), Chatterjee and Dethlefs (2020), Silva et al. (2019), Baghestani and AbuAl-Foul (2019), Silva et al. (2017), and Hassani et al. (2017).

5. Forecasting Results

We now turn to the forecasting performance of the SDM methods when applied to real data. We consider forecasting horizons of 1, 3, 6 and 12 months ahead and compare the forecasts obtained by SDM with the forecasts produced by classical time series models, such as ARIMA, ETS, TVP, non-linear STAR models and with a machine learning method, Neural Networks (NN). Using the data in the form of annual growth, the forecasting performance was tested, setting the first 14 years of data as training window (168 observations) and the last 8 years (96 observations) as the post-sample period. Like the simulation studies, here we also used the initial stretch of data (say m observations) and fitted a linear AR model to produce initial estimates of the parameters, residual variance, variance-covariance matrix, and the recursion is then started from midway along the initial stretch of data at $t = m/2$. To obtain the forecasts, the window expansion approach has been employed and for each series the recursion started with the same initial values for the parameters. Data from January 2011 were then used as out-of-sample and RMSE and RRMSE were employed to assess the forecasting accuracy for all models.

The KSPA test (Hassani and Silva, 2015) was also conducted to compare forecasting errors between ETS, ANN, ESTAR, LSTAR, TVP and the three types of SDM models against the forecasting errors of the benchmark naïve model. The RMSE based forecasting results are reported in the supplementary data through Tables A1-A4 whilst the RRMSE results are reported in text via Tables 2a-2d. Summary statistics in the tables below provide the average Relative Root Mean Squares, and Score shows the number of times (out of 10) that a competing model performed better than the naïve model. The tables also report the number of times that a model performs significantly better than the naïve model (at the 5 or 1% level based on the KSPA test). In brief, the results in Tables A1-A4 show that no single model can provide the best forecast for all series across all horizons.

We begin by analysing the forecasting results for 1-month ahead forecasts. The SDM based forecasts are seen reporting lowest RMSE (Table A1) when forecasting 'Europe Others', 'North America Business', 'North America Others', and the 'Grand Total' at this horizon. Interestingly, the naïve method is seen to be the best model at forecasting the 'Europe Tourism' and 'North America Tourism' series. The RRMSE results in Table 2a compare the findings from all models in relation to the benchmark naïve model. The score indicates that TVP is the best performer based on the RMSE in relation to the naïve model. However, only one of those scores are statistically significant. Likewise, even though the score and sig. score indicate that SDM models have the highest combinations, the significant scores are once again very low, ranging from 3 to 4 out of 10. We find evidence for statistically significant differences between forecasts from the models reporting the lowest RMSE and naïve forecasts in the case of 'Europe Others', 'North America Business', and 'North America Others' only. Nevertheless, in terms of the 'Asia Others' series, we find the ETS forecasts significantly outperform forecasts from the benchmark model whilst for 'Europe Business' two of the SDM based forecasts outperform the naïve forecasts. In general, we found no evidence of statistically significant differences between the naïve forecast and any of the competing models at forecasting total Japanese tourist arrivals at 1-month ahead, thereby indicating that practitioners can even rely on the naïve model at this horizon as opposed to relying on any other complex forecasting models. Overall, the RMSE based findings show that in the very short run no single model can provide the best forecast for all series. However, based on the average RRMSE we can conclude that the classical recursive SDM method is slightly better than other competing methods for forecasting Japanese tourist arrivals at 1-month ahead.

Table 2a: RRMSE of 1-month ahead forecasts.

Sector	ARIMA	ETS	NN	SDM Recursive	SDM Unsmoothed	SDM LOWESS	STAR	LSTAR	TVP
Asia Business	0.941	1.009	**0.818**	0.926	0.918	0.910	0.933	0.958	0.916
Asia Tourism	1.025***	1.089	1.430	1.019	1.039	1.029	1.039	1.042	**0.991**
Asia Others	0.804	0.806*	0.867	0.803	0.808	0.805	0.795	**0.756**	0.811
Europe Business	1.008	1.050	1.076**	0.983	0.980**	0.986**	0.979	0.988	**0.953**
Europe Tourism	**1.080**	1.179	1.259	1.084*	1.238	1.112	1.159	1.118	1.095
Europe Others	0.808	0.793	0.808	0.823*	0.818	**0.792***	0.837	0.812	0.840
North America Business	0.927	1.043	0.905	0.899	0.904	**0.898***	1.063	1.043	0.946
North America Tourism	**1.049**	1.092	1.539	1.019	1.012	1.010	1.045	1.056	1.066
North America Others	0.767**	0.735**	0.739**	**0.724*****	0.781**	0.783**	0.778**	0.765**	0.768**
Grand Total	1.012	1.058	1.211	**0.994**	1.004	0.994	1.010	1.010	0.994
Summary									
Average RRMSE	0.942	0.985	1.065	**0.927**	0.950	0.932	0.964	0.955	0.938
Score	5	3	5	7	6	7	5	5	8
Sig. Score	2	2	2	3	2	4	1	1	1

Note: ****/**/* indicates a statistically significant difference between the benchmark naïve forecast and a competing forecast based on the KSPA test at p=0.01, p=0.05 and p=0.10 respectively. Highlighted in bold font is the model reporting the lowest RMSE for a particular time series.*

Next, we consider the forecasting results at 3-months ahead. Based on the RMSE, Table A2 indicates that as the forecasting horizon increases beyond the very short run, the accuracy of SDM models appear to improve beyond previous levels. In fact, at this horizon, the SDM models are seen providing the lowest RMSE for 7 out of the 10 time series analysed in this study. Interestingly, the SDM-LOWESS model provides the lowest RMSE at forecasting aggregated Japanese tourist arrivals at the 3-months ahead horizon too. Table 2b indicates that 7/10 times the model reporting the lowest RMSE fails to produce forecasts which are statistically significantly better than the benchmark naïve model. However, the RRMSE results for 'Asia Business', 'Asia Tourism', 'Asia Others', 'Europe Business', and 'Europe Tourism',

indicate there is a competing model that performs significantly better than the naïve model at generating forecasts for these series at $h = 3$ months-ahead. The naïve model is recommended for 'Europe Others', and the 'Grand Total'. The score shows that forecasts from all models except ETS and NNETAR outperform the naïve forecasts always based on the lower RMSE. However, again, the significance scores are very low except in the case of SDM Recursive forecasts where 60% of the scores are statistically significant. Accordingly, it appears that as the horizon increases beyond the very short run, the need for more complex models increases. Overall, based on the average RRMSE (Table 2b), the two new forecasting methods outperform the other competing methods and beginning to perform comparatively better than they did at $h = 1$ month-ahead. The Recursive SDM also outperforms other methods.

Table 2b: RRMSE of 3-month ahead forecasts.

Sector	ARIMA	ETS	NN	SDM Recursive	SDM Unsmoothed	SDM LOWESS	STAR	LSTAR	TVP
Asia Business	0.790	0.935	**0.7217**	0.771**	0.756*	0.754*	0.7789**	0.849	0.776*
Asia Tourism	0.891***	0.985	1.940	0.856	0.856	**0.853**	0.870	0.930	0.910
Asia Others	0.780	0.790**	0.828	0.767	0.756	**0.746**	0.774	0.775	0.850
Europe Business	0.826	0.940	1.162	0.790***	**0.768**	0.773	0.788	0.770	0.874
Europe Tourism	0.976	1.012	1.171	0.903**	**0.882**	0.894	0.981	0.941	0.974
Europe Others	0.801	0.790	0.795	0.806	0.799	0.808	0.806	**0.772**	0.812
North America Business	0.756*	0.946	0.729	0.734**	**0.722****	0.722**	0.800	0.935	0.823*
North America Tourism	0.968	0.991	2.601	**0.879****	0.884	0.884	0.904	0.959	0.908
North America Others	0.827*	**0.776***	0.819**	0.847**	0.858	0.855	0.847*	0.800**	0.866
Grand Total	0.930	0.995	1.126*	0.899	0.899	**0.887**	0.912	0.917	0.927
Summary									
Average RRMSE	0.854	0.916	1.189	0.825	**0.818**	**0.818**	0.846	0.865	0.872
Score	10	9	5	10	10	10	10	10	10
Sig. Score	3	2	2	6	2	2	2	1	2

Note: ****/**/* indicates a statistically significant difference between the benchmark naïve forecast and a competing forecast based on the KSPA test at p=0.01, p=0.05 and p=0.10 respectively. Highlighted in bold font is the model reporting the lowest RMSE for a particular time series.*

At the 6-months ahead horizon, Table A3 indicates that forecasts based on SDM report the lowest RMSE for 8 out of the 10 series evaluated here. Interestingly, the NNETAR model continues to provide the lowest RMSE at forecasting 'Asia Business' at horizons of 1, 3 and 6-months ahead. Table 2c shows that the naïve model is best for forecasting Japanese tourist arrivals series 'Europe Others' and 'North America Others' at this horizon. Once again, there are only two instances where the models reporting the lowest RMSEs also report statistically significantly better forecasts than those from the naïve model. However, in comparison to the forecasts at $h = 1$ and $h = 3$ months-ahead, we notice a higher number of cases where there is a competing forecast which significantly outperforms the benchmark naïve forecast with 7/10 cases. Interestingly, in terms of forecasting the 'Grand Total' we see that forecasts from ARIMA significantly outperform the benchmark naïve model. This is the first instance where the benchmark model was significantly outperformed at forecasting aggregated Japanese tourist arrivals. Again, we notice how the usefulness of more complex models is seen to be more prevalent as the forecasting horizon widens further and SDM LOWESS produced the best average RRMSE, as shown in Table 2c.

Table 2c: RRMSE of 6-month ahead forecasts.

Sector	ARIMA	ETS	NN	SDM Recursive	SDM Unsmoothed	SDM LOWESS	STAR	LSTAR	TVP
Asia Business	0.669	0.9205	**0.5369**	0.614*	0.603**	0.601**	0.638*	0.788	0.637*
Asia Tourism	0.771***	0.975	1.975***	0.665	**0.652**	0.653	0.715	0.790	0.800**
Asia Others	0.719**	0.713	0.794	0.703	0.701	**0.660**	0.730**	0.738***	0.742*
Europe Business	0.783*	0.919	1.127	0.655***	0.632***	**0.630*****	0.660***	0.647***	0.763***
Europe Tourism	0.960	1.013	1.107	0.731**	0.686	**0.684**	0.738	0.752	0.854
Europe Others	0.831	0.830	0.802	0.802	0.790	**0.783**	0.815	0.806	0.834
North America Business	0.626**	0.936	0.617	0.623**	0.613***	**0.607*****	0.693	1.091	0.707**
North America Tourism	0.916*	0.987	2.429	0.675	**0.650**	0.655	0.741	0.805	0.759

North America Others	0.783	0.773	0.776	0.784	0.791	0.795	0.806	**0.769**	0.815
Grand Total	0.832**	0.990	1.515*	0.721	0.717	**0.691**	0.792	0.769	0.832
Summary									
Average RRMSE	0.789	0.906	1.168	0.697	0.684	**0.676**	0.733	0.795	0.774
Score	10	9	5	10	10	10	10	9	10
Sig. Score	6	0	2	4	3	3	3	2	5

Note: ****/**/* indicates a statistically significant difference between the benchmark naïve forecast and a competing forecast based on the KSPA test at p=0.01, p=0.05 and p=0.10 respectively. Highlighted in bold font is the model reporting the lowest RMSE for a particular time series.*

At the 12-month ahead horizon, Table A4 indicates that forecasts from techniques based on SDM are reporting the lowest RMSE for all cases except for the 'North America Others' series. It is known that generating accurate forecasts in the very long run is comparatively more difficult than at shorter time horizons. Accordingly, it appears that SDM based forecasts can perform very well at the very long horizon than shorter horizons. Furthermore, forecasts from NNETAR no longer report the lowest RMSE for 'Asia Business' at this horizon. According to Table 2d, this is also the first instance where the naïve model fails to provide a significantly more accurate forecast for any of the Japanese tourist arrivals series. The score indicates that all three SDM models and the STAR model outperforms the benchmark model in all cases and the SDM models report a significance score of 9/10 with the only instance where the SDM forecasts fail to significantly outperform the benchmark model being the case of 'Asia Others'. These represent the highest significance scores reported by any model within this forecasting exercise thereby confirming the usefulness of a mix of SDM based models for long run forecasting of Japanese tourist arrivals in general. Nevertheless, it is noteworthy that for the 'Asia Others' series, the best forecasting model at this horizon would be TVP whereas for the 'North America Others' series the best forecasting model would be LSTAR. Finally, in terms of forecasting the 'Grand Total' at $h = 12$ months-ahead the results evidence the superiority of the SDM LOWESS model in relation to the benchmark. Table 2d also shows that the two new proposed SDMs produced the best average RRMSE.

Table 2d: RRMSE of 12-month ahead forecasts.

Sector	ARIMA	ETS	NN	SDM Recursive	SDM Unsmoothed	SDM LOWESS	STAR	LSTAR	TVP
Asia Business	0.547	0.877	0.487	0.472**	0.472***	**0.471*****	0.509**	0.617	0.530**

Asia Tourism	0.600***	0.943	1.943***	0.471**	**0.446***	0.450***	0.553	0.626	0.709
Asia Others	0.608***	0.614	0.622	**0.562**	0.565	0.582	0.600***	0.613***	0.570***
Europe Business	0.748***	0.859	1.136***	0.516***	**0.485***	**0.485***	0.523***	0.567***	0.604***
Europe Tourism	0.781	0.932	0.979	0.449***	0.425**	**0.413***	0.481	0.506	0.834
Europe Others	0.748	0.745	0.639	0.584*	**0.557****	0.561**	0.604	0.627	0.680
North America Business	0.491**	0.913	0.480	0.485***	0.477***	**0.474***	0.537*	1.443	0.634**
North America Tourism	0.887**	0.957	2.185	0.446*	**0.408****	0.409**	0.559	0.666	0.585
North America Others	0.621**	0.686	0.784	0.640**	0.620**	0.619**	0.644**	**0.589***	0.683**
Grand Total	0.691	0.974	1.618***	0.503*	0.503**	**0.466***	0.628	0.633	0.713
Summary									
Average RRMSE	0.672	0.850	1.087	0.513	0.496	**0.493**	0.564	0.689	0.654
Score	10	10	6	10	10	10	10	9	10
Sig. Score	6	0	3	9	9	9	5	3	5

Note: ****/**/* indicates a statistically significant difference between the benchmark naïve forecast and a competing forecast based on the KSPA test at p=0.01, p=0.05 and p=0.10 respectively. Highlighted in bold font is the model reporting the lowest RMSE for a particular time series.*

Finally, we present Table 3 to summarise and provide readers with a snapshot of the core findings from our study. This table will allow readers to pick the most suitable model for forecasting disaggregated or aggregated Japanese tourist arrivals at a horizon of interest. The score based results in this table, which shows the number of times a model was selected as best for forecasting a time series at a particular horizon, indicates that as the horizon increases, the usefulness of the naïve model diminishes. However, the fact that it is best in 50% of the cases at $h = 1$ month-ahead is a clear indicator of the importance of considering benchmark models as opposed to assuming more complex models lead to better forecasts across all horizons or instances. Furthermore, we see that the SDM Recursive model is not suitable for forecasting any of the series at $h = 12$ months-ahead. Instead, we are able to propose a mix of SDM UNSMOOTHED and SDM LOWESS models as best for forecasting majority of Japanese tourist arrivals series in the very long run.

Table 3: Summary of best model for forecasting aggregated and disaggregated Japanese tourism arrivals.

Series	1-month Ahead	3-months Ahead	6-months Ahead	12-months Ahead
Asia Business	NAÏVE	SDM LOWESS	SDM LOWESS	SDM LOWESS
Asia Tourism	NAÏVE	ARIMA	ARIMA	SDM UNSMOOTHED
Asia Others	ETS	ETS	ARIMA	TVP
Europe Business	SDM UNSMOOTHED	SDM RECURSIVE	SDM LOWESS	SDM LOWESS / SDM UNSMOOTHED
Europe Tourism	NAÏVE	SDM RECURSIVE	SDM RECURSIVE	SDM LOWESS
Europe Others	SDM LOWESS	NAÏVE	NAÏVE	SDM UNSMOOTHED
North America Business	SDM LOWESS	SDM UNSMOOTHED	SDM LOWESS	SDM LOWESS
North America Tourism	NAÏVE	SDM RECURSIVE	ARIMA	SDM UNSMOOTHED
North America Others	SDM RECURSIVE	ETS	NAÏVE	LSTAR
Grand Total	NAÏVE	NAÏVE	ARIMA	SDM LOWESS
Model	**Score**	**Score**	**Score**	**Score**
Naïve	5	2	2	0
ARIMA	0	1	4	0
ETS	1	2	0	0
TVP	0	0	0	1
LSTAR	0	0	0	1
SDM RECURSIVE	1	3	1	0
SDM UNSMOOTHED	1	1	0	4
SDM LOWESS	2	1	3	5
SDM MODELS MIX	4	5	4	8

Note: The best model is the one which reports the lower out-of-sample forecasting RRMSE and statistically significant differences in forecasts in comparison to forecasts from the benchmark naïve model. SDM MODELS MIX refers to the performance of a combination of SDM-based models taken as a whole.

5.1 Comparison between SDM against ARIMA with dummies

To reflect the influence of external shocks, dummy variables may be used. In this section, we compare the forecasting performance of SDM against an ARIMA model with dummy variables. To assess the forecasting accuracy of the two methods in the presence of structural breaks, we

consider the 'Grand Total' series. Japan suffered from the effect of Global Financial Crisis and credit crunch in 2008 and earthquake & tsunami in 2011. To include these two major shocks in our modelling, we extended the in-sample period to include 192 months (16 years), leaving 72 months (6 years) of post sample data. The Bai and Perron (2003) structural breaks test were employed to identify the timing of the shocks and the forecasting performance of SDM were then compared with an AR model with added dummies. .

Table 4 presents the out-of-sample forecasting results for the 'Grand Total' enabling comparison between SDMs and an AR model with dummies. Based on the RMSE, the SDM forecasts are more accurate than the AR model across all horizons. We find evidence of statistically significant differences between the forecasts obtained by AR model with dummies and the SDM LOWES forecasts at $h = 3,6$ and 12 months-ahead. These results once again show the superiority of the SDM LOWES forecasting in the long run.

Table 4: RMSE and RRMSE forecasting results for 'Grand Total' in the presence of external shocks.

Model	1 step	3 steps	6 steps	12 steps
SDM Recursive	7.682	9.272	12.393	14.340
SDM Unsmoothed	8.030	8.533	11.550	14.759
SDM LOWESS	7.183	7.229	9.725	14.023
ARIMA with 2 dummies	8.399	11.536	16.535	21.507
RRMSE Recursive	0.915	0.804	0.749**	0.667***
RRMSE Unsmoothed	0.956	0.740	0.698**	0.686***
RRMSE LOWESS	0.855	0.627**	0.588***	0.652***

Note:*/**/* indicates a statistically significant difference between forecasts from an ARIMA model with 2 dummy variables and the SDM forecast based on the KSPA test at p=0.01, p=0.05 and p=0.10 respectively.*

6. Concluding remarks and future research

In this paper, we expand the tools available for forecasting tourism demand by successfully introducing two new forecasting methods based on State-Dependent Models. The estimation performance of SDMs are first evaluated via a Monte Carlo simulation study and the forecasting performance of the SDMs are assessed through an application to real data.

The Japanese tourism data examined in this research experienced structural breaks and non-linearity. Linear time series models do not usually account for these characteristics and

therefore may provide inaccurate forecasts. This research investigated the effectiveness of the estimation technique of a general non-linear model, State-Dependent Model. The study also examined and evaluated the forecasting performance of this general non-linear model relative to the naïve model, ARIMA, ETS, Neural network models, TVP and STAR models using Japanese Tourism data. We contributed to assessing the forecasting performance of two newly proposed SDM approaches. The three SDM forecasting methods along with other linear and nonlinear time series models such as: ARIMA, ETS, and Neural Network, STAR and TVP models were used to forecast the Grand Total of inbound Japanese tourist arrivals and their components and compared with the naïve forecasts.

The results indicated the superiority of the SDM models compared to the competing methods in some cases. The improvements became more pronounced at longer time horizons. The newly proposed SDM forecasting method generally performed better than the traditional recursive SDM forecasting model for long-term forecasting. The study also sought to understand why the SDM forecasts can achieve superior forecasting results compared to the other competing models. Applying Monte Carlo simulations, we generated data from various linear/non-linear models to examine the effectiveness of the SDM estimation technique. The results were encouraging, showing that the SDM algorithm can correctly estimate the true functional form of parameters based on the data only, without any priori knowledge priori of the underlying model, and clearly distinguished different types of models. We also generated time series with a structural break, with a degree of shift in the level. Based on the results presented in the simulation section, we showed that the intercept term μ can capture the dynamics of the movements in data and shifts in the level of the series. The time plot of $\hat{\mu}$ for the real data (please see the Appendix) employed in the study showed that the SDM could capture sudden changes in the mean. Graphs of all the $\widehat{\phi}$ parameters against both time and state for the 10 series are also given in the Appendix. Further Monte Carlo simulation results for different models are also available from the authors upon request.

Overall, the simulation results in this study showed the effectiveness of the SDM estimation method by applying it on data generated from various time series models. It also showed the superiority of the SDM forecasting methods over the classical times series and machine learning approaches, especially for long term forecasting. The Monte Carlo simulation study indicated that the SDM estimation technique works well and can estimate the coefficients accurately, pointing us clearly in the direction of the true model without any priori knowledge of the underlying model. The ability that a model fit better in-sample is likely to produce better out-of-sample forecasts as proven by the empirical findings when applied to real data.

The findings also show that the two proposed SDM forecasting methods can improve forecast errors compared to the traditional SDM forecasting method in the long-term horizon. However, in terms of RRMSE, the results indicated no significant difference between the two proposed forecasting methods.

Through this study we have successfully introduced a new approach for tourism demand forecasting, adding to the variety of options available, thereby giving forecasters and practitioners more choice. We have also shown that the SDM can handle structural breaks and produce more accurate forecasts at the longer forecast horizons. This is crucial as external shocks have significant impact on the tourism industry. Overall, SDMs combined with LOWESS can be a reliable model for Japanese tourism demand forecasting (Table 3) at longer horizons. The comparatively better performance of SDM-based models at longer horizons is useful for practitioners as the generation of accurate forecasts become difficult in the long run. Our research also provides several benefits for tourism management because better medium-term and long-term forecast accuracy translates into better allocation of resources for tourist amenities, hotels, staffing, attractions, and airlines. Furthermore, governments can anticipate tourist arrivals up to one year with greater confidence thereby enabling more efficient resource allocation. The Japanese government can also make better decisions on efforts on promoting Japanese tourism and provide better support to the airline and hotel industry through effective budget proposals provided the long-term forecasts are accurate. Suggestions for future research and further developments in SDM are listed below:

a) Adopt a time-varying residual variance, σ_ϵ^2 , in the algorithm, allowing it to vary from one point to the next. This updating can be done, for example, by using the information on the current values of the residuals so far computed. This method adds a heteroscedastic variance into the model.

b) Instead of setting and using a constant smoothing parameter for all time periods in the SDM algorithm, a time varying smoothing parameter with the most appropriate value from one time point to the next, may be used so that problem of outliers in the time series data may be dealt more effectively.

c) Assess the forecasting performance of the State Dependent approach in a further Monte Carlo simulation study, generating data from various non-linear models, where the functional form of the parameters depends on different lags.

d) Due to data unavailability, we have not considered the impact of COVID-19 on SDM models. Future studies should consider evaluating the sensitiveness of SDM models against the pandemic.

References

Álvarez-Díaz, M. and Rosselló-Nadal, J. 2010. Forecasting British tourist arrivals in the Balearic Islands using meteorological Variables. *Tourism Economics* 16(1), pp. 153-168. doi: 10.5367/000000010790872079

Armstrong, G. W. G. 1972. International tourism: coming or going. *Futures* 4(2), pp. 115-125. doi: 10.1016/0016-3287(72)90036-5

Athanasopoulos, G. and de Silva, A. 2012. Multivariate exponential smoothing for forecasting tourist arrivals. *Journal of Travel Research* 51(5), pp. 640-652. doi: 10.1177/0047287511434115

Baghestani, H. and AbuAl-Foul, B. 2019. Predicting United Arab Emirates' real effective exchange rates using oil prices. *OPEC Energy Review* 43(4), pp. 492-511.

Bai, J. and Perron, P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18(1), pp. 1-22. doi: 10.1002/jae.659

Bergmeir, C., Hyndman, Rob J. and Benítez, José M. 2016. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting* 32(2), pp. 303-312. doi: 10.1016/j.ijforecast.2015.07.002

Bi, J.-W., Li, H. and Fan, Z.-P. 2021. Tourism demand forecasting with time series imaging: A deep learning model. *Annals of Tourism Research* 90, doi: 10.1016/j.annals.2021.103255

Bi, J.-W., Liu, Y. and Li, H. 2020. Daily tourism volume forecasting for tourist attractions. *Annals of Tourism Research* 83, doi: 10.1016/j.annals.2020.102923

Cartwright, P. A. 1985a. Forecasting time series: a comparative analysis of alternative classes of time series models. *Journal of Time Series Analysis* 6(4), pp. 203-211.

Cartwright, P. A. 1985b. Using state dependent models for prediction of time series with missing observations. *Time Series Analysis: Theory and Practice 7*, pp. 157-167.

Cartwright, P. A. and Newbold, P. 1982. *A time series approach to the prediction of oil discoveries.*

Chatterjee, J. and Dethlefs, N. 2020. Deep learning with knowledge transfer for explanable anomaly prediction in wind turbines. *Wind Energy* 23, pp. 1693-1710.

Chen, C.-F., Chang, Y.-H. and Chang, Y.-W. 2009. Seasonal ARIMA forecasting of inbound air travel arrivals to Taiwan. *Transportmetrica* 5(2), pp. 125-140. doi: 10.1080/18128600802591210

Cho, V. 2016. Tourism forecasting and its relationship with leading economic indicators. *Journal of Hospitality & Tourism Research* 25(4), pp. 399-420. doi: 10.1177/109634800102500404

Cleveland, W. S. 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74(368), pp. 829-836. doi: 10.1080/01621459.1979.10481038

Fan, G.-F., Jin, X.-R. and Hong, W.-C. 2021. Application of COEMD-S-SVR model in tourism demand forecasting and economic behavior analysis: The case of Sanya City. *Journal of the Operational Research Society*, pp. 1-13.

Goh, C. and Law, R. 2002. Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management* 23(5), pp. 499-510. doi: 10.1016/s0261-5177(02)00009-2

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), pp. 296-317. doi: 10.1080/10548408.2011.562856

Goh, C., Law, R. and H.M., M. 2008. Analyzing and forecasting tourism demand: A rough sets approach. *Journal of Travel Research* 46(3), pp. 327-338.

Haggan, V., Heravi, S. M. and Priestley, M. B. 1984. A study of the application of state-dependent models in non-linear time series analysis. *Journal of Time Series Analysis* 5(2), pp. 69-102.

Hassani, H. and Silva, E. S. 2015. A Kolmogorov-Smirnov based test for comparing the predictive accuracy of two sets of forecasts. *Econometrics* 3(3), pp. 590-609. doi: https://doi.org/10.3390/econometrics3030590

Hassani, H., Silva, E. S., Antonakakis, N., Filis, G. and Gupta, R. 2017. Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research* 63, pp. 112-127. doi: 10.1016/j.annals.2017.01.008

Hassani, H., Webster, A., Silva, E. S. and Heravi, S. 2015. Forecasting U.S. tourist arrivals using optimal Singular Spectrum Analysis. *Tourism Management* 46, pp. 322-335. doi: 10.1016/j.tourman.2014.07.004

He, K., Ji, L., Wu, C. W. D. and Tso, K. F. G. 2021. Using SARIMA–CNN–LSTM approach to forecast daily tourism demand. *Journal of Hospitality and Tourism Management* 49, pp. 25-33.

Huang, B. and Hao, H. 2021. A novel two-step procedure for tourism demand forecasting. *Current Issues in Tourism* 24(9), pp. 1199-1210.

Law, R. 2000. Back-propagation learning in improving the accuracy of neural network-based tourism demand forecasting. *Tourism Management* 21(4), pp. 331-340. doi: 10.1016/s0261-5177(99)00067-9

Law, R., Li, G., Fong, D. K. C. and X., H. 2019. Tourism demand forecasting: A deep learning approach. *Annals of Tourism Research* 75, pp. 410-423.

Ozaki, T. 1981. Non-linear threshold autoregressive models for non-linear random vibrations. *Journal of Applied Probability* 18, pp. 443-451.

Priestley, M. 1980a. Prediction based on a general class of non-linear models. *Technical Report No. 126, Department of Mathematics, UMIST*,

Priestley, M. B. 1980b. State-dependent models: a general approach to non-linear time series analysis. *Journal of Time Series Analysis* 1(1), pp. 47-71. doi: 10.1111/j.1467-9892.1980.tb00300.x

Qiu, R. T. R., Wu, D. C., Dropsy, V., Petit, S., Pratt, S. and Ohe, Y. 2021. Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and Pacific team. *Annals of Tourism Research* 88, doi: 10.1016/j.annals.2021.103155

Radenović, F., Tolias, G. and Chum, O. 2018. Fine-tuning CNN image retrieval with no human annotation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41(7), pp. 1655-1668.

Sheldon, P. J. 1993. Forecasting tourism: expenditure versus arrivals. *Journal of Travel Research* 32, pp. 13-20.

Silva, E. S., Ghodsi, Z., Ghodsi, M., Heravi, S. and Hassani, H. 2017. Cross country relations in European tourist arrivals. *Annals of Tourism Research* 63, pp. 151-168.

Silva, E. S., Hassani, H., Heravi, S. and Huang, X. 2019. Forecasting tourism demand with denoised neural networks. *Annals of Tourism Research* 74, pp. 134-154. doi: 10.1016/j.annals.2018.11.006

Song, H., Qiu, R. T. R. and Park, J. 2019. A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research* 75, pp. 338-362. doi: 10.1016/j.annals.2018.12.001

Tian, X., Wang, H. and E, E. 2021. Forecasting intermittent demand for inventory management by retailers: A new approach. *Journal of Retailing and Consumer Services* 102662,

UNWTO. 2019. *International tourism highlights: 2019 edition.*

Witt, S. F., Newbould, G. D. and Watkins, A. J. 2016. Forecastin domestic tourism demand: application to Las Vegas arrivals data. *Journal of Travel Research* 31(1), pp. 36-41. doi: 10.1177/004728759203100108

World Travel & Tourism Council. 2020. *Japan 2020 annual research: key highlights*. Available at: https://wttc.org/Research/Economic-Impact [Accessed: 25 August 2021].

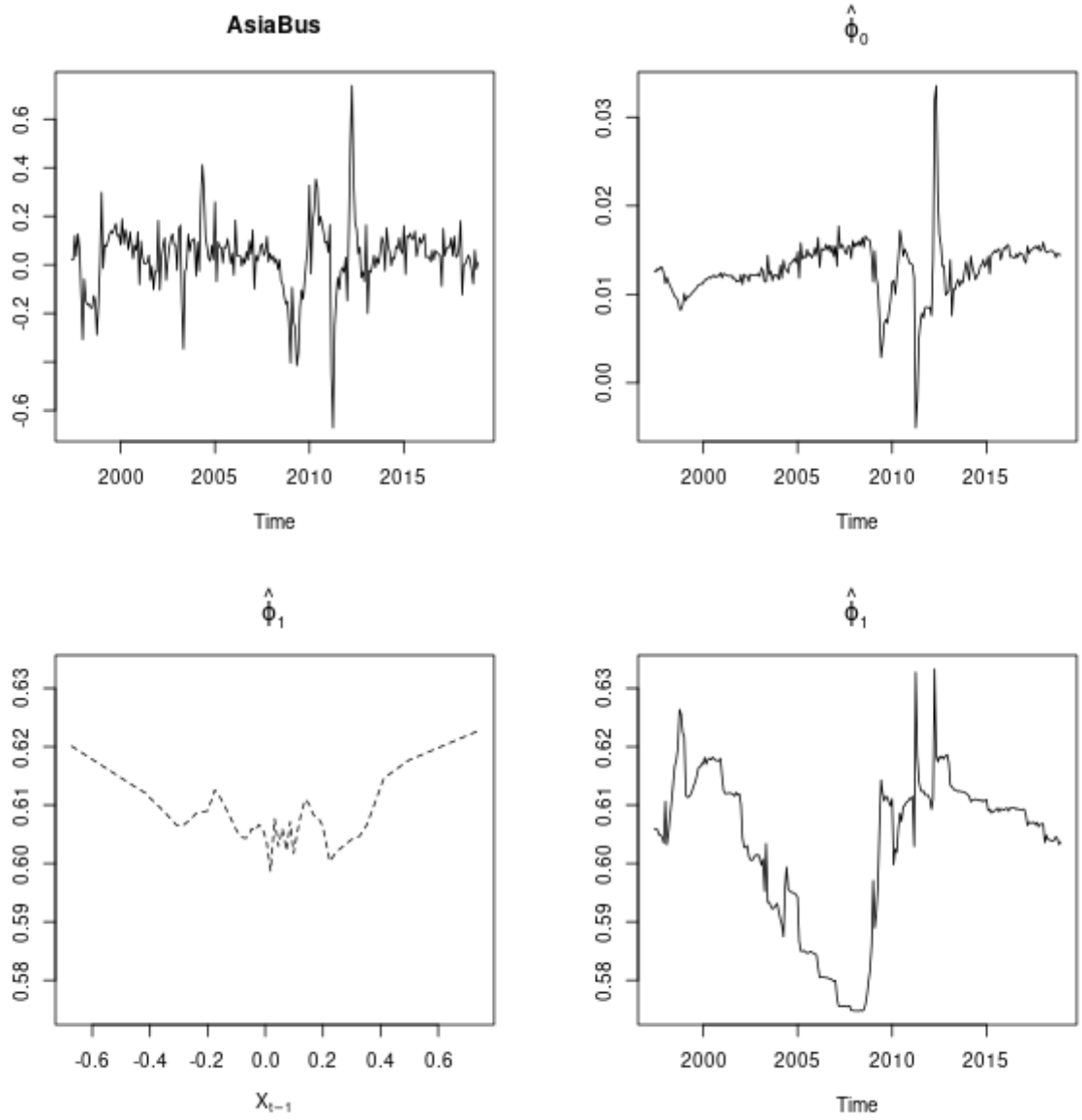
Zhang, H., Song, H., Wen, L. and Liu, C. 2021. Forecasting tourism recovery amid COVID-19. *Annals of Tourism Research* 87, pp. 103-149. doi: 10.1016/j.annals.2021.103149

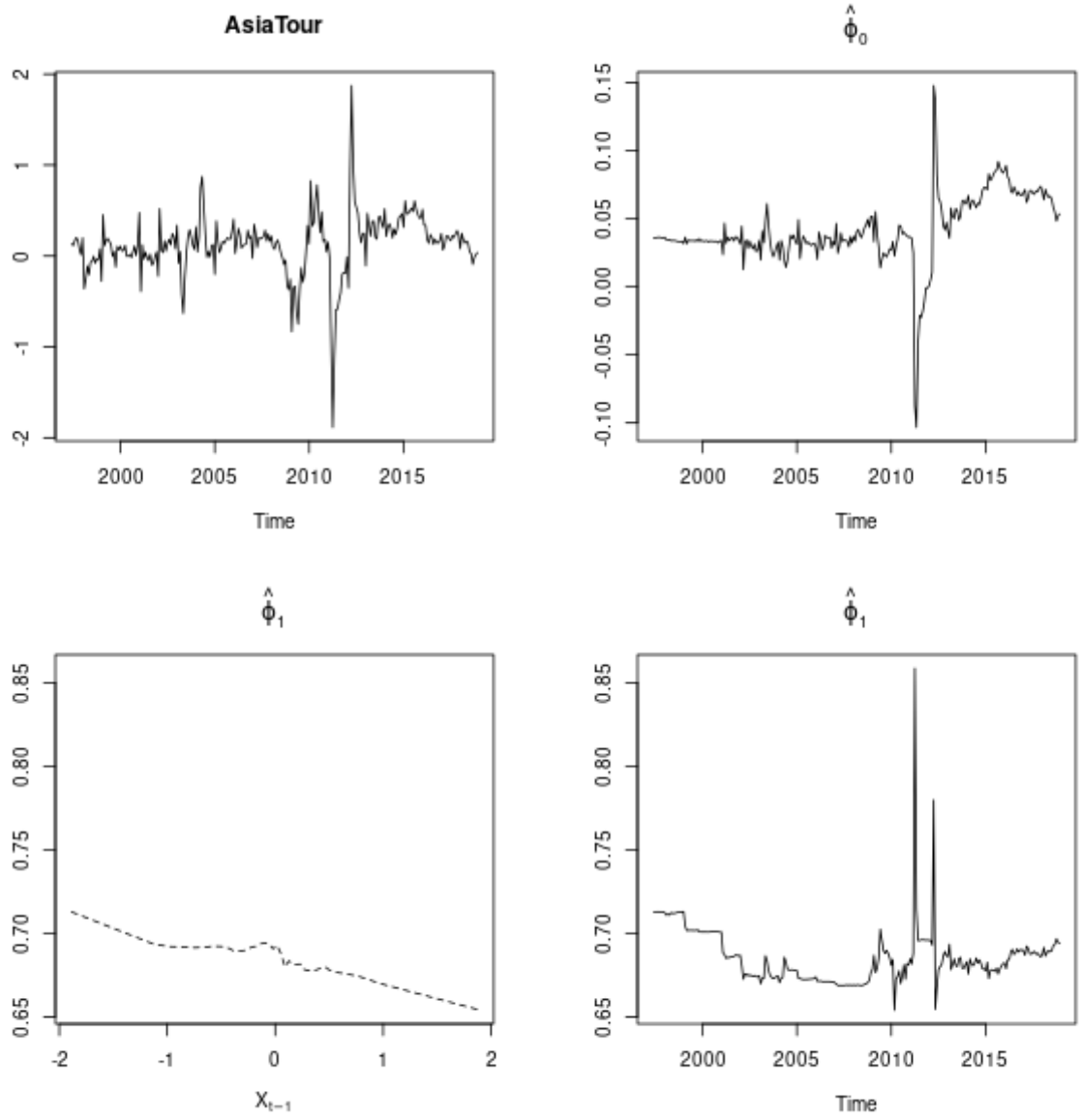
Zhang, K., Zuo, W., Chen, Y., Meng, D. and Zhang, L. 2017. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process* 26(7), pp. 3142-3155. doi: 10.1109/TIP.2017.2662206

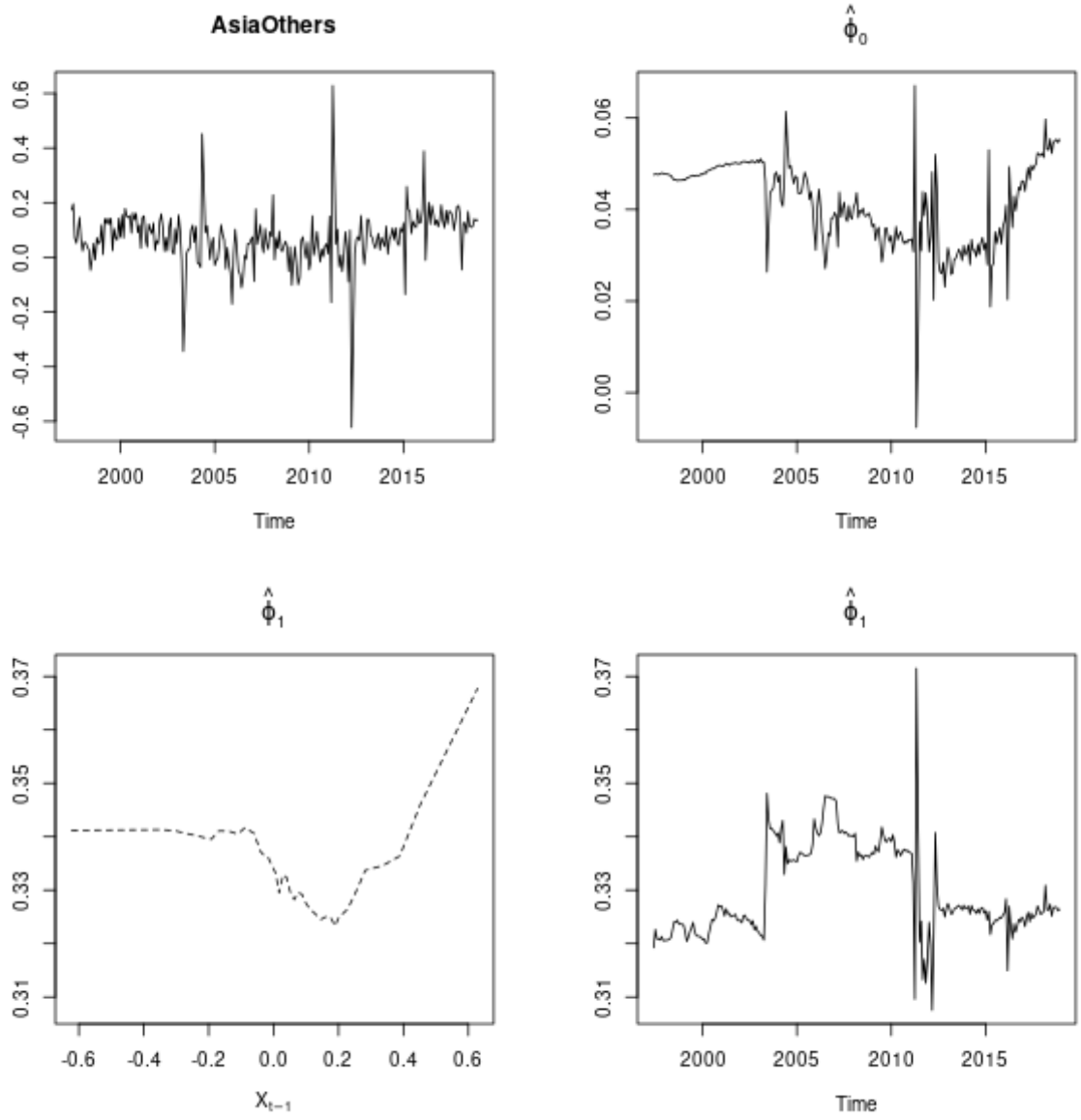
Zheng, W., Huang, L. and Lin, Z. 2021. Multi-attraction, hourly tourism demand forecasting. *Annals of Tourism Research* 90, p. 103271. doi: 10.1016/j.annals.2021.103271

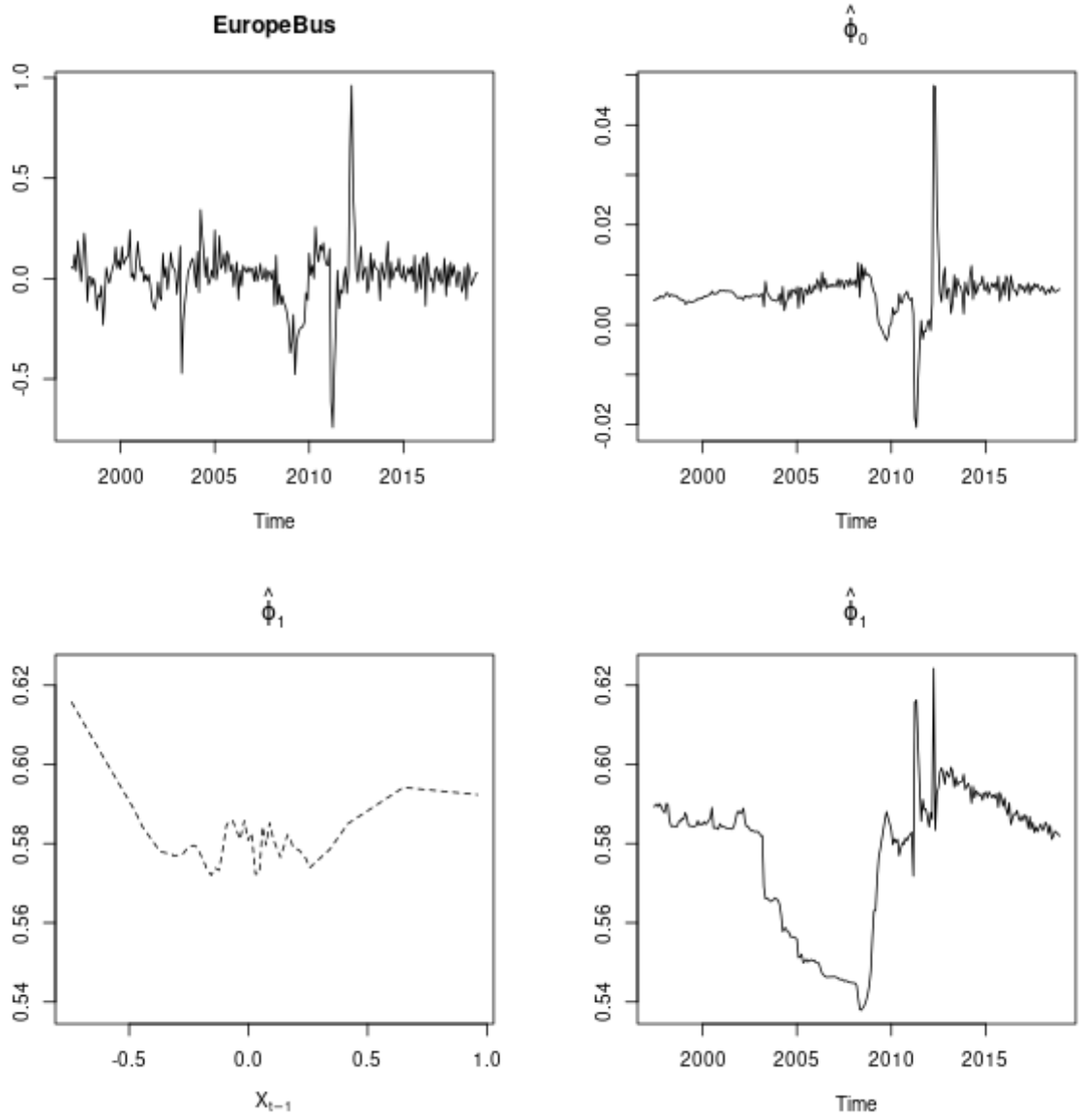
Appendix:

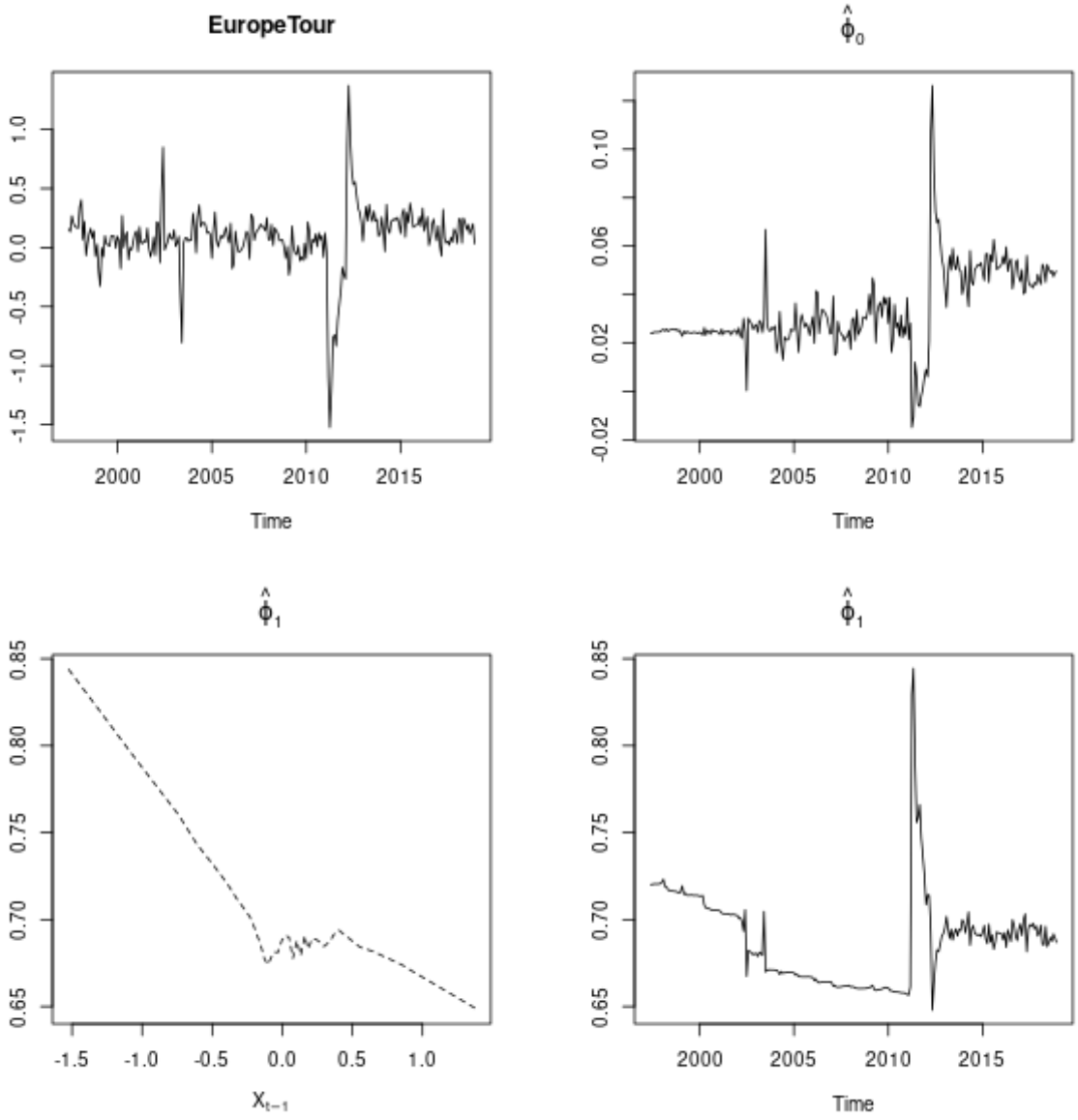
Plots of original times series, $\hat{\phi}_0$ against time & $\hat{\phi}_1$ against both X_{t-1} and time for the 10 series

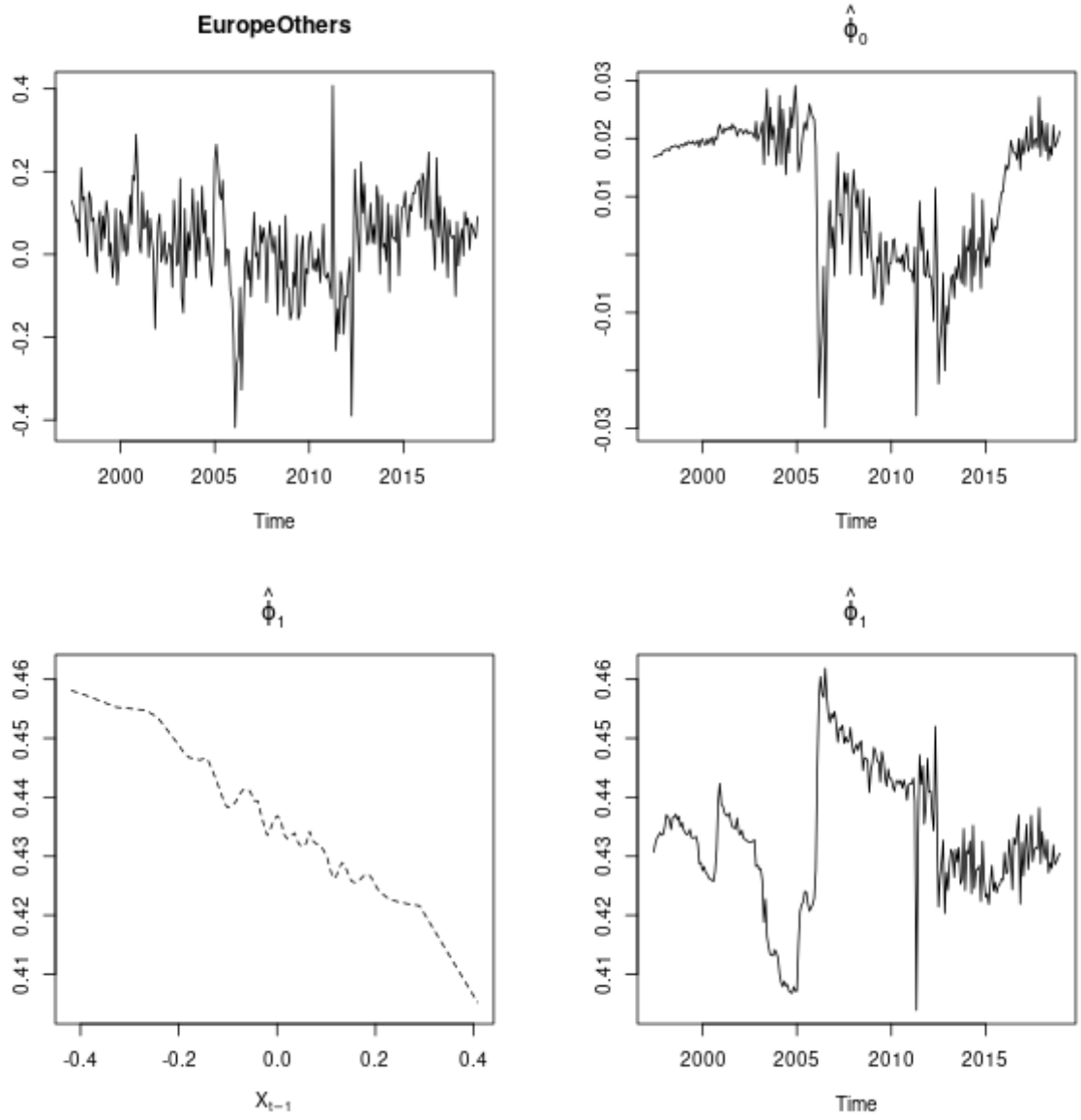


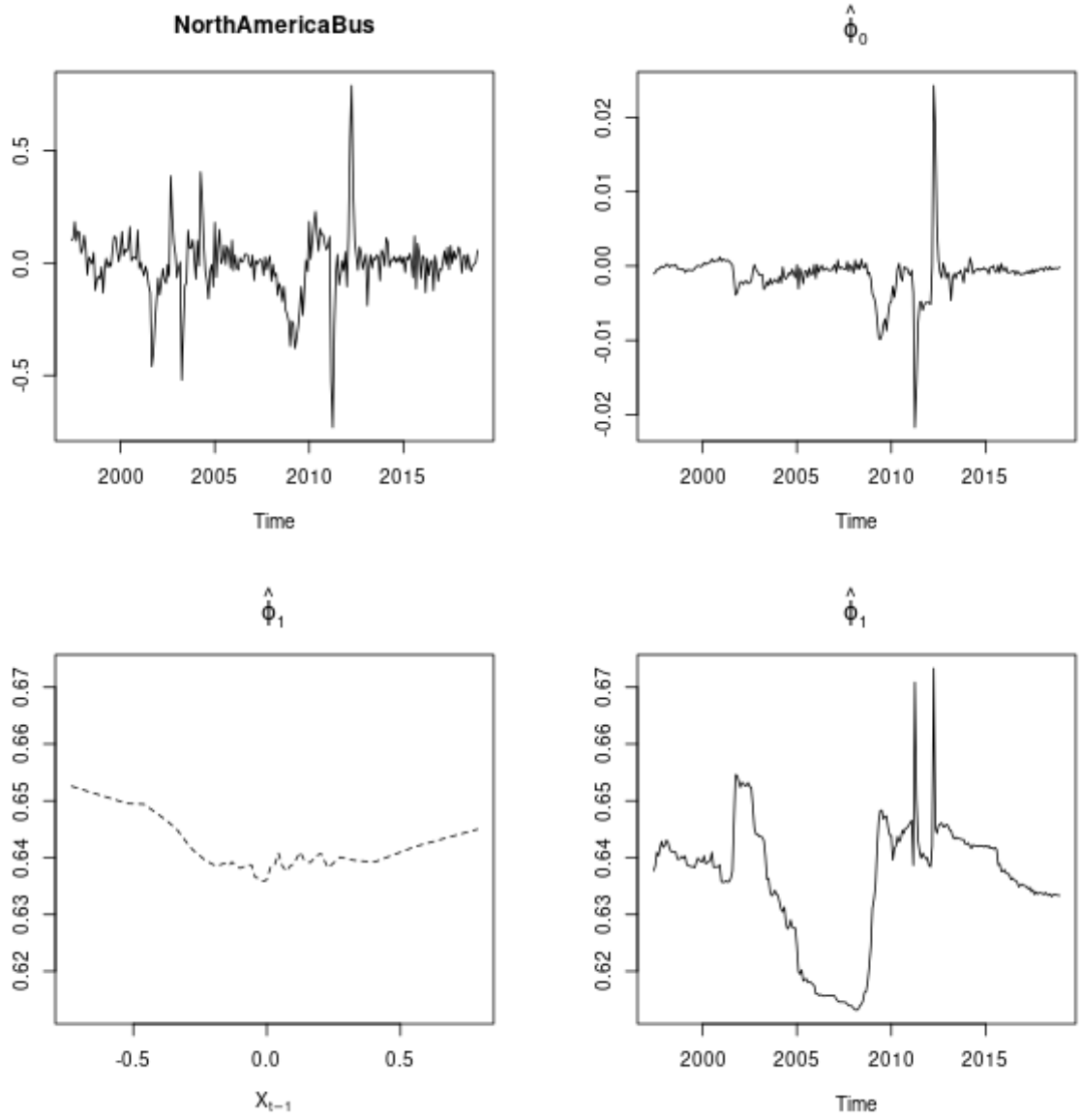

AsiaTour
2
1
0
-1
-2
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.15
0.10
0.05
0.00
-0.05
-0.10
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.85
0.80
0.75
0.70
0.65
-2
-1
0
1
2
X_{t-1}
$\hat{\phi}_1$
0.85
0.80
0.75
0.70
0.65
2000
2005
2010
2015
Time

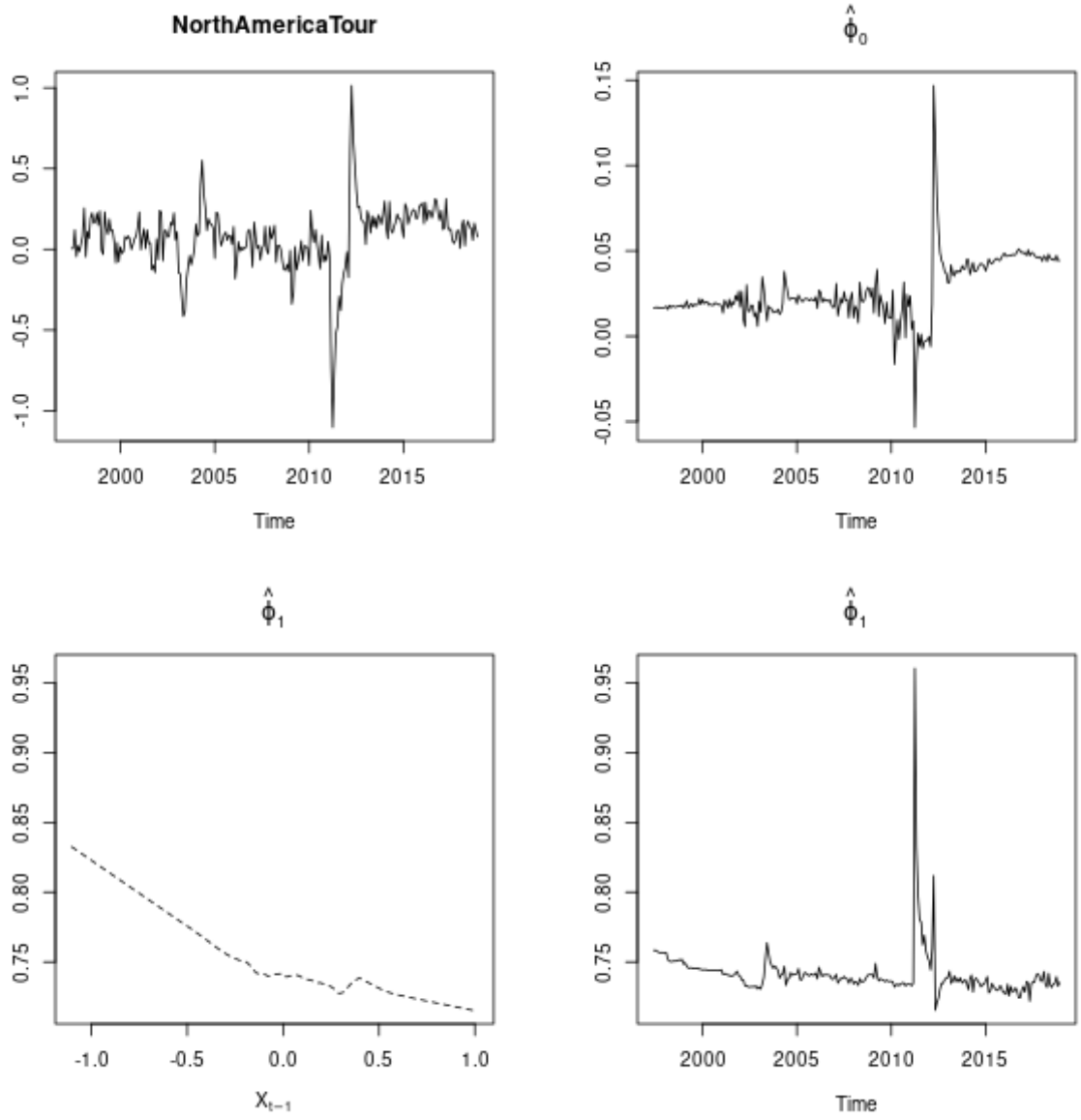

AsiaOthers
Time
$\hat{\phi}_0$
Time
$\hat{\phi}_1$
X_{t-1}
$\hat{\phi}_1$
Time

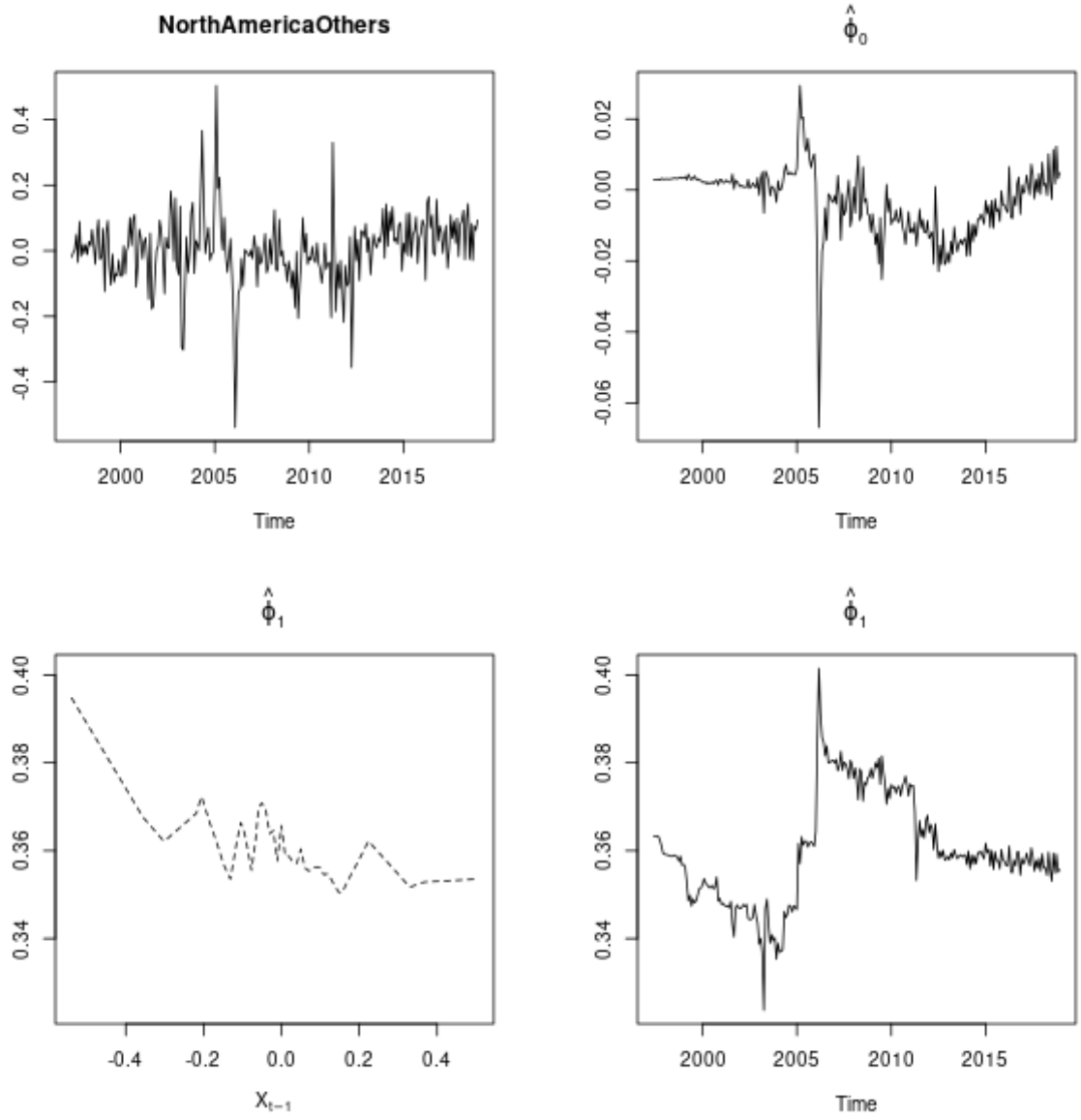

EuropeBus
1.0
0.5
0.0
-0.5
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.04
0.02
0.00
-0.02
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.62
0.60
0.58
0.56
0.54
-0.5
0.0
0.5
1.0
X_{t-1}
$\hat{\phi}_1$
0.62
0.60
0.58
0.56
0.54
2000
2005
2010
2015
Time

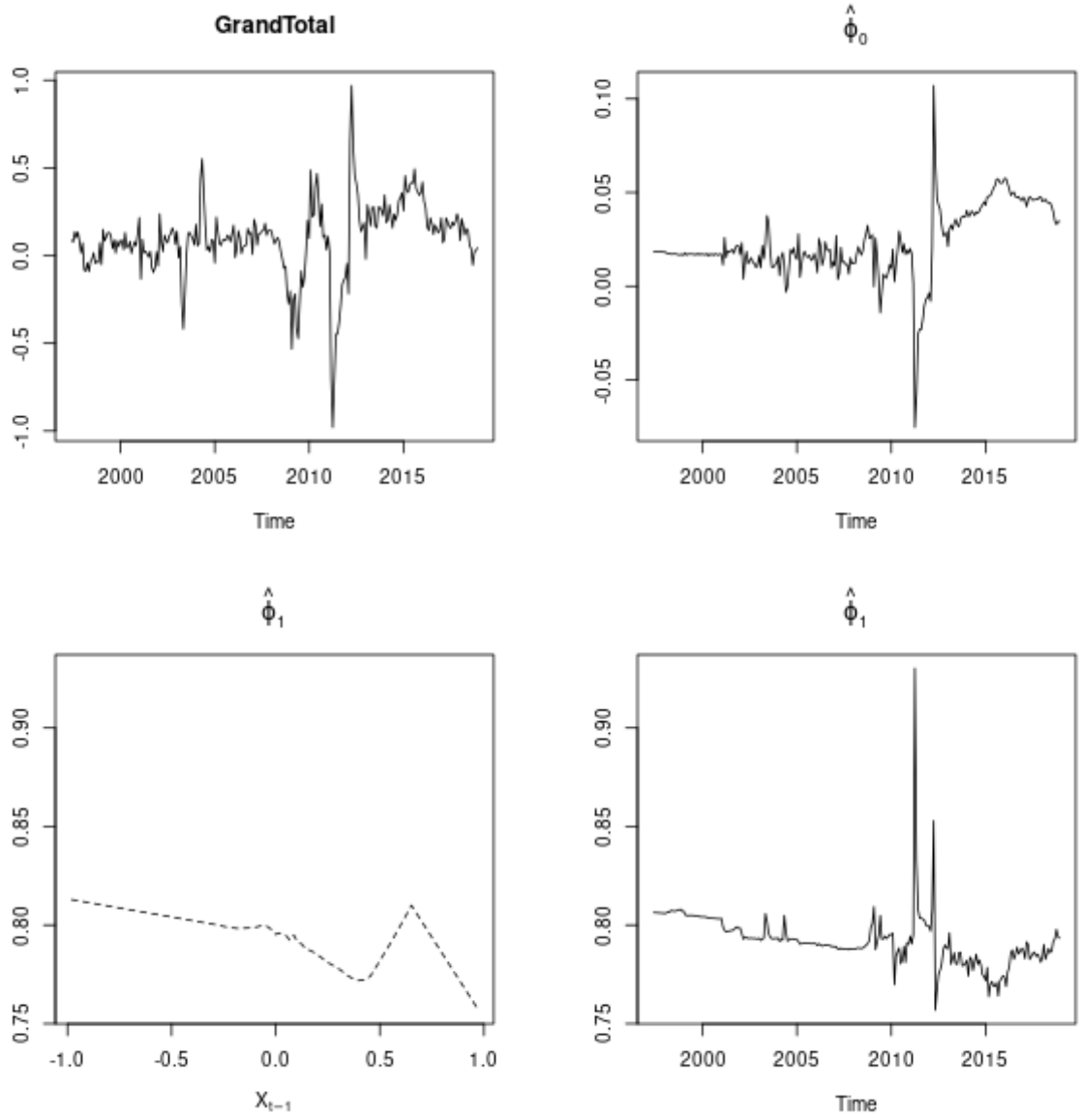

EuropeTour
Time
$\hat{\phi}_0$
Time
$\hat{\phi}_1$
X_{t-1}
$\hat{\phi}_1$
Time


EuropeOthers
0.4
0.2
0.0
-0.2
-0.4
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.03
0.02
0.01
-0.01
-0.03
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.46
0.45
0.44
0.43
0.42
0.41
-0.4
-0.2
0.0
0.2
0.4
X_{t-1}
$\hat{\phi}_1$
0.46
0.45
0.44
0.43
0.42
0.41
2000
2005
2010
2015
Time


NorthAmericaBus
0.5
0.0
-0.5
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.02
0.01
0.00
-0.01
-0.02
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.67
0.66
0.65
0.64
0.63
0.62
-0.5
0.0
0.5
X_{t-1}
$\hat{\phi}_1$
0.67
0.66
0.65
0.64
0.63
0.62
2000
2005
2010
2015
Time


NorthAmericaTour
1.0
0.5
0.0
-0.5
-1.0
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.15
0.10
0.05
0.00
-0.05
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.95
0.90
0.85
0.80
0.75
-1.0
-0.5
0.0
0.5
1.0
X_{t-1}
$\hat{\phi}_1$
0.95
0.90
0.85
0.80
0.75
2000
2005
2010
2015
Time


NorthAmericaOthers
0.4
0.2
0.0
-0.2
-0.4
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.02
0.00
-0.02
-0.04
-0.06
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.40
0.38
0.36
0.34
-0.4
-0.2
0.0
0.2
0.4
X_{t-1}
$\hat{\phi}_1$
0.40
0.38
0.36
0.34
2000
2005
2010
2015
Time


GrandTotal
1.0
0.5
0.0
-0.5
-1.0
2000
2005
2010
2015
Time
$\hat{\phi}_0$
0.10
0.05
0.00
-0.05
2000
2005
2010
2015
Time
$\hat{\phi}_1$
0.90
0.85
0.80
0.75
-1.0
-0.5
0.0
0.5
1.0
X_{t-1}
$\hat{\phi}_1$
0.90
0.85
0.80
0.75
2000
2005
2010
2015
Time