International Journal of Forecasting 19 (2003) 435–451

international journal of forecasting

www.elsevier.com/locate/ijforecast

Univariate versus multivariate time series forecasting: an application to international tourism demand

Johann du Preez[a], Stephen F. Witt[b,c,*]

[a]ACNielsen, Oxford OX3 9RX, UK

[b]School of Management, University of Surrey, Guildford GU2 7XH, UK

[c]Centre for Regional and Tourism Research, Bornholm, Denmark

Abstract

Tourist numbers from several origin countries to a particular destination country form a vector series. In the presence of a 'rich' cross-correlation structure, that is if after allowing for autocorrelation the sample cross-correlation function exhibits meaningful and statistically significant correlations, the accuracy when forecasting a particular origin–destination tourist flow is likely to be improved by utilising information from the other tourist flows. Multivariate time series models may be expected to generate more accurate forecasts than univariate models in this setting. However, in the absence of these conditions, univariate forecasting models may well outperform multivariate models. An empirical investigation of tourism demand from four European countries to the Seychelles shows an absence of such a 'rich' structure and that ARIMA exhibits better forecasting performance than univariate and multivariate state space modelling. One implication that an absence of a 'rich' cross-correlation structure holds for econometric modelling is that explanatory variables which are strongly correlated with the tourist flow series are likely to be uncorrelated across origin countries.

Keywords: Comparative methods; Time series; Univariate; Multivariate; ARIMA; State space; Tourism forecasting

1. Introduction

Within the set of variables that an organisation forecasts, there are likely to be groups of variables that follow similar time-based patterns. Various phenomena can give rise to such similarities. For example, when a country's national tourist organisation examines time series on international tourism demand from several generating countries to that destination, it may be the case that these time series are influenced in a similar manner by global forces. In particular, tourism demand from different origin countries located in the same general area and operating under similar economic, political and social conditions is likely to be affected in the same sort of manner by economic and environmental factors (e.g. regional booms/recessions and changing consumer tastes, and wars/terrorist activities/civil unrest). The covariation of time series that follow similar time-based patterns is a source of information that can improve forecast accuracy in certain situations.

Accurate forecasts of tourism demand are essential for efficient planning by the various sectors of the

*Corresponding author. Tel.: +44-1483-686-320; fax: +44-1483-686-301.

E-mail address: a.west@surrey.ac.uk (S.F. Witt).

0169-2070/02/$ – see front matter
doi:10.1016/S0169-2070(02)00057-2

tourism industry, and forecast accuracy is particularly important in the tourism context as the tourism product is perishable (e.g. unused aeroplane seats, hotel rooms and hire car rentals cannot be stockpiled). Specifically, short-term forecasts can aid decision making in areas such as scheduling, staffing and planning tour operator brochures.

Many studies of international tourism demand analyse tourist flows from several individual origin countries to a particular destination country. Recent examples which employ time series methods include Turner, Kulendran and Pergat (1995) (inbound tourism to New Zealand broken down into 11 source countries/regions); Kulendran and King (1997) (inbound tourism to Australia from four origin countries); Turner, Kulendran and Fernando (1997) (inbound tourism to Japan, Australia and New Zealand, each from four origin countries); and Kim and Song (1998) (inbound tourism to South Korea from four origin countries). In each of these studies several types of model are estimated, but all the time series models are univariate; e.g. ARIMA models are used to generate forecasts in each case, but only univariate ARIMA models are considered. Thus each tourism flow is examined in isolation; no information from one such series is used in estimating a different tourist flow. When forecasting international tourism demand information of this nature may, however, improve forecasting accuracy. An investigation of the potential usefulness of multivariate time series models for forecasting within a tourism context, where the multivariate nature of the models lies in the cross-correlations of 'parallel' series, should therefore be illuminating. Such multivariate forecasting models do not appear to have been applied previously to tourism.

There are several tourism forecasting studies which compare the accuracy of short-term forecasts generated by multivariate models incorporating explanatory variables with the accuracy of forecasts generated by univariate models. The extra complexity of multivariate tourism forecasting models does not necessarily lead to an improvement in performance. For example, the error correction model has been shown to be outperformed by the 'no change' (random walk) model (Song & Witt, 2000; Kulendran & Witt, 2001); and by the univariate ARIMA model (González & Moral, 1995; Kulendran & King, 1997). Furthermore, the causal structural time series model has been shown to be outperformed by the univariate ARIMA model (Garcia-Ferrer & Queralt, 1997). On the other hand, it has been demonstrated that the Bayes hierarchical model forecasts tourism demand more accurately than the 'no change' model (Stroud, Sykes & Witt, 1998); and the time varying parameter model generates more accurate forecasts than the 'no change' and univariate ARIMA models (Song & Witt, 2000).

In the general forecasting literature the empirical evidence on univariate versus multivariate models is also conflicting. For example, Brodie and de Kluyver (1987) have shown that naïve model forecasting will often yield more accurate results than econometric models when predicting market share. By contrast, Allen and Fildes (2001) discovered, when examining studies that compared econometric and extrapolative forecasts, that in a majority of cases the econometric forecasts were more accurate. In addition, Bidarkota (1998) has shown that the error correction model generates more accurate forecasts than a univariate unobserved components model.

The methodology for building univariate and multivariate (in terms of cross-correlations of 'parallel' series) time series models is well known and has become fairly automatic. A vast array of user-friendly software enables practitioners to obtain forecasts based on various models with ease. Unfortunately, the apparent ease and simplicity of the process may be quite misleading. For instance, the availability of good multivariate state space modelling software and the observation that the tourist numbers form a vector series may tempt practitioners into making multivariate state space modelling the preferred and only forecasting approach. It is intuitively appealing that developing a multivariate model where the parameters are estimated by minimising the determinant of the error (or residual) covariance matrix will result in a model where the determinant of the prediction error covariance matrix of the vector series is small compared to that of other models. Although the trace is usually preferred to the determinant of the prediction covariance matrix as a measure of multivariate forecasting performance (primarily for the greater analytical tractability of the former), both are valid measures of multivariate variance. Again the availability of (say) state space modelling software may clinch the decision as to which model to use in view of the fact that state space models belong to the multivariate class of models where the parameters are estimated by minimising the determinant of the error covariance matrix. In some circumstances such a decision may turn out to be correct but it could equally well be a poor decision. According to current good practice, even when interest is focused on multivariate models, such as ARIMA or transfer function models, univariate models are first fitted to the components of the multiple time series. The cross-correlations between the observed residuals of the component series (or alternatively prewhitened component series) play a pivotal role in any subsequent analysis. The reason for this is quite simply that it is not a trivial matter to decide whether two stationary component series are uncorrelated, unless one of the component series is white noise (Box & Jenkins, 1976; Granger & Newbold, 1977). In this setting the hypothesis that the series are uncorrelated is therefore plausible only if satisfactory univariate models for the component series are available, and the component series can be prewhitened or alternatively the observed residual series of the components used. In this study the second approach is adopted, but all results are also valid for the former approach.

On an intuitive level, one would expect that if the sample cross-correlation structure between component series of a multiple time series is weak, that is it exhibits few significant cross-correlations (that may well have appeared purely by chance), state space models, transfer function models and univariate ARIMA models should, in principle, produce fairly similar results with respect to univariate and multivariate short term forecasting criteria. In practice, models involving a greater number of parameters may be at a disadvantage, and parameter rich models such as state space models at a particular disadvantage. This study provides an example where, in dealing with a vector series, a set of univariate models with parameters estimated so as to minimise univariate error variance yields a smaller determinant of the short term prediction error matrix than the multivariate state space modelling approach, even though the parameter estimation in the latter is based on minimising the determinant of the covariance matrix of residuals. In fact, in this example the short term forecasting behaviour of the set of univariate models is generally superior to that obtained from state space models. In other situations, particularly in the presence of a rich sample cross-correlation structure between the prewhitened component time series, any inadequacy arising from parameterisation and estimation problems should be more than compensated for by exploiting the structure. To summarise, it may well be conjectured that in the presence of a ‘weak’ cross-correlation structure, univariate forecasting models exhibit better forecasting performance in terms of both univariate and multivariate criteria than either multivariate state space or transfer function models. In the presence of a richer cross-correlation structure, transfer function models should exhibit better univariate criteria forecasting performance and multivariate state space models better multivariate criteria forecasting performance.

The data set examined in this study comprises monthly observations on visitor arrivals in the Seychelles from its major European origin countries: France, Germany, Italy and the UK. The visitor arrivals data are published by the Management and Information Systems Division of the Seychelles Ministry of Administration and Manpower in *Statistical Bulletin: Tourism* and are obtained from immigration cards completed by all arriving visitors. The data cover the period January 1980 to December 1994 (180 observations for each of the four time series)[1].

The following models are considered:

1. Univariate ARIMA models for the four component series.
2. Univariate state space models for the four component series[2]. The models are developed using two different estimation methods, two different approaches to determining state space dimension and two different methods of deleting insignificant parameters.
3. Multivariate state space models (and by implication multivariate ARMA models) for the vector of the four component series.

[1] These data are available from the authors upon request.

[2] It is not suggested that state space modelling should be used for forecasting univariate time series. The aim is to highlight features of the procedure and/or algorithm used in the software.

In each case, the models are estimated using data for the period 1980–92, thus retaining the data for 1993–94 to evaluate the external validity of the models obtained. Throughout the present study SAS System software, Release 6.12, is used for all aspects of the data analysis. Although the use of different software may lead to slight differences in model estimation, it should not affect the overall conclusions to any significant degree.

The purposes of the paper are as follows:

1. To compare the fit and forecasting performance of the multivariate and univariate models.
2. To select the 'best' of the available adequate models, on the basis of the above comparison, for forecasting purposes.
3. To compare the fitted models across the four component series in order to determine whether the structure of the series differ, and therefore whether there are inherent differences among the tourism series from the four generating countries.

We wish to draw inferences, if suggested or justified by the foregoing analysis, relevant to the fields of forecasting practice on the one hand and tourism demand on the other.

The finer detail of most of the modelling is omitted so as not to obscure the primary objectives of the study. This detail is however available on request from the authors.

2. Data transformations

A single outlier appeared in the data for French tourists to the Seychelles for February 1990[3]. It had no real effect on the analysis and consequently has been replaced by an estimate so as not to be distracting in graphical presentations. The timeplots for the series for the four origin countries appear in Fig. 1. The series are clearly non-stationary, exhibiting trends, seasonal components and variance instability (the latter confirmed by plots of standard deviations against means of the monthly tourist numbers). Transformation of the variables involved was required to stabilize the variances; logarithmic transformations seem to work for the UK, German and French series, but not for the Italian series where a power (square root) transformation appears to be more appropriate. The latter has been scaled to be roughly of the same order as the other series.

The working series were transformed by differencing the series (regular and/or seasonal) to eliminate trend, where the order of the differencing required was determined by considering timeplots, sample autocorrelation functions, variances and Dickey–Fuller type tests of the differenced series. These indicators suggested taking one regular difference and one seasonal difference (strongly indicated for the UK and Italian series, marginally less clear-cut as far as the French and German series are concerned). If we denote x_{ci} for observation i (that is month i) from the original series c, where $c = F, G, U$ or I signifies French, German, UK or Italian, respectively, B for the backward shift operator (that is, $Bx_{ci} = x_{c,i-1}$) and y_{ci} for observation i from the transformed series c, then:

$$y_{ci} = (1 - B)(1 - B^{12}) \ln x_{ci} \text{ for } c = F, G, U \text{ and all } i$$

and

$$y_{ci} = (1 - B)(1 - B^{12})(0.1\sqrt{x_{ci}}) \text{ for } c = I \text{ and all } i.$$

Inspection of the cross-covariances between these series confirmed that the vector series $\{\mathbf{Y_i}\}$ where $\mathbf{Y_i}^{\mathrm{T}} = (y_{Fi}, y_{Gi}, y_{Ui}, y_{Ii})$ may be considered a second order stationary vector series.

3. Univariate ARIMA model

The standard procedure for identification, estimation, diagnostic checking and overfitting in a Box–Jenkins analysis of time series was performed. The estimation method involved maximum likelihood parameter estimation to obtain initial estimates and then unconditional least-squares estimation to obtain final estimates (exact least squares method). The results obtained were very similar to those obtained

[3]The arrival of a cruise ship carrying French passengers increased arrivals of French residents in the Seychelles for that month to 4446. Using cubic splines the figure has been reduced to 2514.

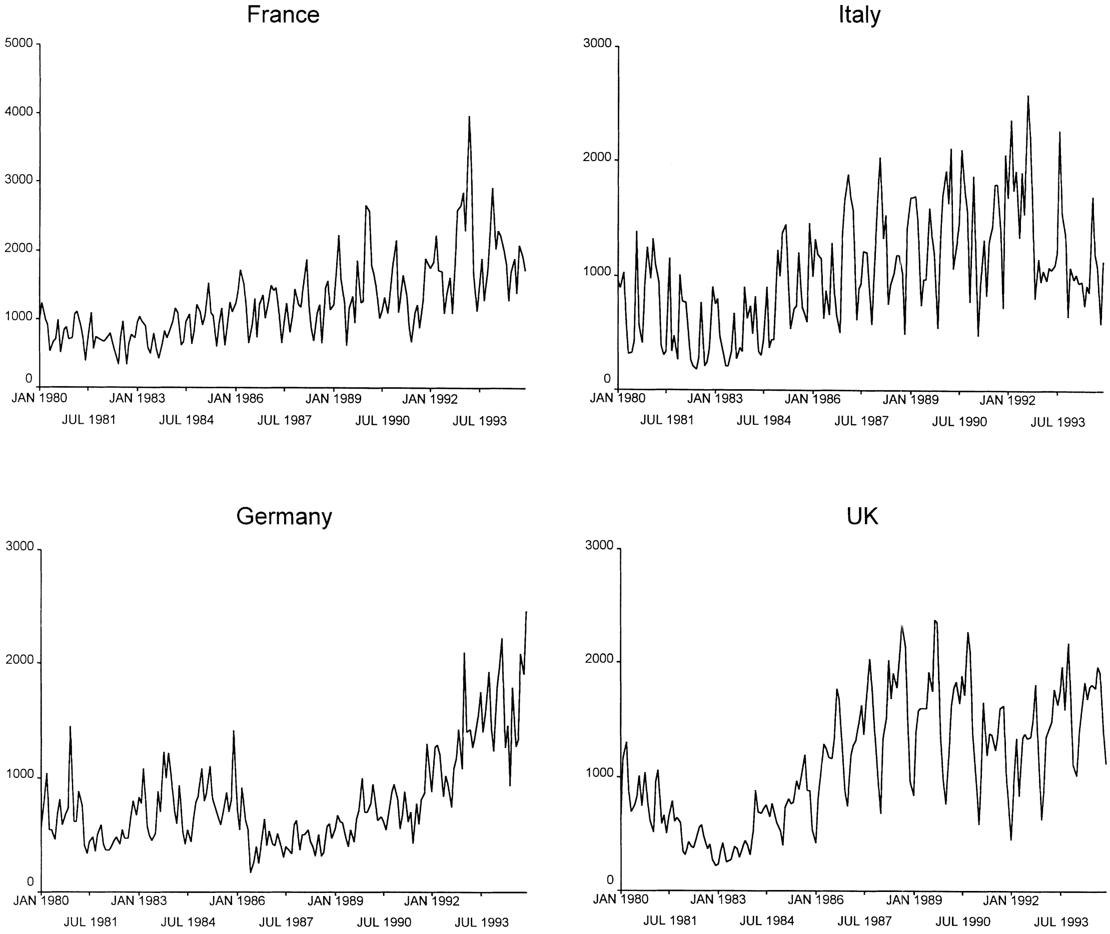

Fig. 1. Monthly tourist numbers.

using maximum likelihood or conditional least squares methods. Model selection was as usual a compromise between goodness of fit and the principle of parsimonious parameterisation. The criteria used for selection were: (a) minimum residual variance for a fixed number of parameters; and (b) a decrease in residual variance of no less than 1% at the cost of an increase in the number of parameters, provided there is a decrease in an information criterion such as Schwarz's Bayesian criterion. The diagnostic results show that the choice of model is not unique, but in all four cases an ARIMA $(0,1,0)\times(1,1,1)_{12}$ model provided an adequate description of the transformed series in terms of fit and distribution of residuals (with the exception of the French data where the residuals appear to deviate marginally from a Gaussian white noise model).

As discussed in the Introduction, the sample correlations and, in particular, the sample cross-correlations of the estimated residuals play a pivotal role in the decision whether or not to proceed with multivariate modelling, and in the modelling itself if it is indicated. The observed residuals may be considered a realisation of a white noise process facilitating inference.

The empirical results show that all but one of the lag 0 sample cross-correlations are significant (but small in magnitude) at the 5% level, whereas the significant sample cross-correlations at non-zero lags may be so purely by chance, which in turn suggests that the residual series are uncorrelated at non-zero lags. This seems quite plausible given the nature of the data. The overall impression is of a fairly weak cross-correlation structure, too weak in fact, as this example is intended to show, to justify the substantial increase in parameterisation demanded by moving to a multivariate model.

The above results yield some useful insights as far as the application area (tourism demand) is concerned. The structures of the component models are remarkably similar apart from the stabilising transformation in the case of Italy. The difference in the variability of the Italian series may well be explained by climatological factors. The southern European/Mediterranean climate experienced in Italy differs substantially from the climate experienced by the majority of the populations in the other three countries. This may have a bearing on the variability in tourism demand to a destination with a tropical climate such as the Seychelles. However, apart from the stabilising transformation, the similarity in structure of the various series—similar component series and uncorrelated (at non-zero lags) residual series—suggests that the same factors (up to components removable by second order differencing) influence tourism demand to the Seychelles from the four origins. Furthermore, it appears that these factors are either weakly correlated if they impact strongly on the component series, or, alternatively, if they are strongly correlated then their unlagged explanatory contributions are relatively weak.

Standard deviations of the error (fit) series appear in Table 1, together with two multivariate measures of fit derived from the matrix of observed error covariances, namely the trace (total variation) and the determinant (generalised variance). Measures of the external validity of the models employed appear in Table 2. A more comprehensive list of univariate measures of forecasting performance for some models is provided in Table 9 (see Garcia-Ferrer, Del Hoyo & Martin-Arroyo, 1997), to be used as a basis for evaluating differences in forecasting performance between the different classes of forecasting models. In order to investigate the effect on forecasting performance of using simpler models, two alternative sets of ARIMA models are used; in the first set the Italian 'best fit' model is replaced by a $(0,1,0)\times(1,1,1)_{12}$ model (retaining the German, French and UK 'best fit' models); and in the second set the German and French 'best fit' models are additionally replaced by a $(0,1,0)\times(1,1,1)_{12}$ model. In the latter situation all four univariate models are $(0,1,0)\times(1,1,1)_{12}$.

The 'best fit' ARIMA models consistently overestimate the Italian series and underestimate the UK and German series; however, the forecasting performance of the models employed is acceptable on the whole, with the actual series well within the

Table 1
Measures of fit for ARIMA models (1980–92)

	Univariate measure (Error Standard Deviation)			
	France	Germany	Italy	UK
'Best'	0.19	0.25	0.41	0.18
'Best'/modified Italy	0.19	0.25	0.42	0.17
$(0,1,0)\times(1,1,1)_{12}$	0.20	0.26	0.42	0.18
	Multivariate measures			
	Total MSR	Generalised MSR $\times 10^{-3}$	Total error variance	Generalised error var $\times 10^{-3}$
'Best'	0.30	1.12	0.30	1.13
'Best'/modified Italy	0.31	1.16	0.31	1.17
$(0,1,0)\times(1,1,1)_{12}$	0.31	1.27	0.32	1.28

Notes: MSR = mean uncorrected sum of squares of residuals. Total MSR = trace of matrix of mean cross-products of observed residuals and the determinant of the matrix is the generalised MSR. Similarly, total error variance and generalised error variance are the trace and determinant, respectively, of the observed residual covariance matrix.

Table 2
Measures of forecasting performance for ARIMA models (1993–94)

	Univariate measure			
	France	Germany	Italy	UK
Mean PE				
'Best'	−0.12	0.47	−0.74	0.46
'Best'/modified Italy	−0.12	0.47	−0.22	0.46
$(0,1,0)\times(1,1,1)_{12}$	−0.11	0.36	−0.22	0.46
PE standard deviation				
'Best'	0.20	0.22	0.33	0.17
'Best'/modified Italy	0.20	0.22	0.30	0.17
$(0,1,0)\times(1,1,1)_{12}$	0.20	0.22	0.30	0.17
	Multivariate measures			
	Total PE	Generalised PE$\times 10^{-5}$	Total PE variance	Generalised PE var$\times 10^{-5}$
'Best'	1.22	0.86	0.23	0.55
'Best'/modified Italy	0.70	0.45	0.21	0.41
$(0,1,0)\times(1,1,1)_{12}$	0.60	0.35	0.21	0.38

Notes: PE = prediction error. Total PE and generalised PE are the trace and determinant, respectively, of the matrix of uncorrected PE cross-products. The total and generalised PE variances are the trace and determinant, respectively, of the covariance matrix of forecast error.

Table 3
Measures of fit for univariate state space observed residuals (1980–92)

Model dimension	Parameter estimation and elimination	Univariate measure (Residual standard deviation)			
		France	Germany	Italy	UK
Free	ml, Akaike	0.44	0.27	0.42	0.18
Free	ml, t-ratio	0.31	0.29	0.42	0.17
Free	cls, Akaike	0.19	0.24	0.42	0.17
Free	cls, t-ratio	0.19	0.24	0.42	0.17
Fixed 13	ml, Akaike	0.20	0.27	0.42	0.21
Fixed 13	ml, t-ratio	0.19	0.29	0.42	0.18
Fixed 13	cls, Akaike	0.19	0.24	0.42	0.17
Fixed 13	cls, t-ratio	0.19	0.24	0.42	0.17
		Multivariate measures			
		Generalised MSR$\times 10^{-5}$	Total MSR	Generalised var $\times 10^{-5}$	Total variance
Free	ml, Akaike	7.45	0.47	7.39	0.47
Free	ml, t-ratio	4.01	0.39	4.01	0.39
Free	cls, Akaike	0.98	0.30	0.96	0.30
Free	cls, t-ratio	1.06	0.30	1.04	0.30
Fixed 13	ml, Akaike	1.86	0.33	1.85	0.33
Fixed 13	ml, t-ratio	1.75	0.33	1.74	0.33
Fixed 13	cls, Akaike	0.91	0.29	0.90	0.29
Fixed 13	cls, t-ratio	1.09	0.30	1.07	0.30

Notes: Free = state space vector determined on the basis of AIC. Fixed 13 = state space vector dimension fixed at 13. Akaike and t-ratio refer to the method of deleting non-significant parameters, the former is based on the least AIC contribution and the latter in order of least non-significant t-ratio, parameters deleted only if the AIC difference is negative. ml = maximum likelihood estimation, cls = conditional least squares.

confidence bounds. It appears that if forecasting needs to be improved, attention should focus on the Italian series.

As can be seen from these tables, within the class of ARIMA models it is well worth compromising on fit to achieve better forecasting performance. The simplest models have total and generalised prediction errors smaller than the best fit models by as much as 50%, at a cost of increasing the total and generalised error (fit) by 4% and 15%, respectively (the error variance of individual series increases by at most 4%). It is quite clear that as far as forecasting models are concerned, the $(0,1,0)\times(1,1,1)_{12}$ models are preferable. They are adequate, fit well and their similarity of structure is appealing as far as the application area is concerned. Moreover, the sample cross-correlations of the observed residuals produced by the different models are similar to those produced by the 'best fit' models and the same conclusions can be drawn.

4. Univariate state space models

It is a fairly common occurrence that differencing a time series introduces moving average terms into the resultant ARIMA model, as seen in the previous section. This has certain implications for state space modelling. The latter first determines the order of an AR model to be fitted to the data on the basis of Akaike's information criterion (AIC). This fixes the number of lags into the past to use in a canonical

Table 4
Measures of univariate state space model observed forecast error (1993–94)

Model dimension	Parameter estimation and elimination	Univariate measures (PE mean and standard deviation)			
		France	Germany	Italy	UK
Free	ml, Akaike	−0.34	0.56	0.16	0.52
		1.38	0.29	0.30	0.21
Free	ml, t-ratio	−0.40	0.65	0.16	0.47
		0.39	0.27	0.30	0.20
Free	cls, Akaike	−0.42	0.39	0.25	0.45
		0.38	0.25	0.31	0.21
Free	cls, t-ratio	−0.41	0.41	0.25	0.46
		0.38	0.25	0.31	0.20
Fixed 13	ml, Akaike	−0.32	0.56	0.17	0.43
		0.33	0.29	0.30	0.23
Fixed 13	ml, t-ratio	−0.45	0.65	0.17	0.43
		0.37	0.27	0.30	0.23
Fixed 13	cls, Akaike	−0.30	0.39	0.24	0.43
		0.33	0.25	0.30	0.19
Fixed 13	cls, t-ratio	−0.48	0.41	0.24	0.48
		0.30	0.25	0.30	0.20
		Multivariate measures			
		Total PE	Generalised PE$\times 10^{-4}$	Total error variance	Generalised error var$\times 10^{-5}$
Akaike	ml, Akaike	2.76	41.33	2.13	49.48
Akaike	ml, t-ratio	1.17	3.22	0.36	2.88
Akaike	cls, Akaike	0.92	1.75	0.34	2.28
Akaike	cls, t-ratio	0.94	1.75	0.34	2.25
Fixed 13	ml, Akaike	0.95	1.63	0.33	2.00
Fixed 13	ml, t-ratio	1.31	2.66	0.35	1.48
Fixed 13	cls, Akaike	0.77	1.01	0.30	1.30
Fixed 13	cls, t-ratio	0.96	1.61	0.30	1.39

correlation analysis, which in turn determines an appropriate state space vector dimension. If the model underlying the data is of the moving average type, then a high order AR model would be required to fit the data well. Unfortunately, if data sets are small this may not be possible, particularly if the moving average component is seasonal. As a precaution the univariate models will be fitted in two different ways. In the first, the dimension of the state space model is determined by canonical correlation analysis (Aoki, 1987) and, in the second, the dimension is fixed so as to capture the seasonal moving average component.

Once the form of a state space equation is decided, initial estimates of free parameters are obtained by maximising likelihood (sample autocovariance matrix based). Non-significant parameters can then be deleted in a backwards elimination process, one at a time. There are two commonly used approaches. In the first, parameters with non-significant t-ratios are selected for consideration in increasing t-ratios order. If the parameter under consideration is non-significant (using AIC for instance), the parameter is dropped, other parameters re-estimated and the process continued. In the second approach, parameters with non-significant t-ratios are selected in increasing order of AIC, deleted if found to be non-significant and the process continued. Once the process is completed, the final maximum likelihood (ML) estimates are obtained. This final model is also re-estimated using conditional least squares (CLS) estimation. The ML estimation method employs an algorithm that effectively ignores the sample ACF after a cut-off point to be specified by the user. Given the moving average nature of the limited data available, this algorithm may lead to less stable optimisation results. Estimates based on CLS may be more reliable. It is not that the algorithm for ML estimation is unreliable, but that the nature of the present data may lead to non-convergence of the ML estimates.

To summarise, for each component series eight models are fitted corresponding to free and fixed dimension state space vectors, ML or CLS estimation, and t-ratio or AIC parameter elimination. The measures of fit are summarised in Table 3 and the forecasting performance measures appear in Table 4. More commonly used measures of forecasting performance are presented in the comparative Table 9. The following conclusions concerning fit may be made:

- There is no general evidence that one method of backward elimination is superior to the other, but if CLS estimation is used AIC is uniformly better.
- There is no evidence of superiority of one method of determining state space vector over the other. Forecasting performance is perhaps more relevant in this case.
- CLS estimation yields smaller estimated residual variance (not necessarily mean sums of squared residuals) than ML.

Conclusions concerning forecasting performance are as follows:

- There is no significant difference in forecasting performance between the two methods of non-significant parameter elimination.
- The method of state space dimension determination has hardly any effect on forecasting performance. It seems that any advantage in capturing the seasonal component in the AR approximation is lost in the disadvantage of using more parameters.
- CLS estimation results in substantially better forecasting performance, particularly with regard to multivariate measures.
- If conditional least squares estimation is used, least AIC contribution elimination of parameters provides on balance better forecasting performance.
- The best overall performance is achieved for the fixed state space dimension model with CLS estimation and least AIC elimination of parameters. This model may be considered as ‘best’ within the class of univariate state space models.

Diagnostic checking establishes that the ‘best’ models are adequate, and that in respect of cross-correlations between residual series the same conclusions may be drawn as in ARIMA modelling. The UK and German series forecasts underestimate the actual series, whereas the French forecasts overestimate the observed series. Overall there seems to be an acceptable correspondence between the patterns that can be noticed in the observed series and those of the forecasts (with the exception of the German series).

5. Multivariate models — state space modelling

5.1. Model selection

The methods used in this section are primarily due to Akaike (1976). Variations on these methods are recommended in some situations (Aoki, 1987), but in this study they led to the same conclusions.

The state space representation of a stationary multivariate series $\mathbf{x_t}$ of dimension r is of the form:

$$\mathbf{z_t} = \mathbf{Az_{t-1}} + \mathbf{Ce_t}$$

$$\mathbf{x_t} = \mathbf{Hz_t} = [\mathbf{I_{r\times r}}: \mathbf{0_{r\times s-r}}]\mathbf{z_t}$$

where $\mathbf{z_t}$ is a vector process of dimension s, $s \geq r$; the first r components of $\mathbf{z_t}$, known as the state space vector, compose $\mathbf{x_t}$; $\mathbf{A}$ is a $s \times s$ transition matrix; $\mathbf{C}$ is a $s \times r$ matrix known as the input matrix; $\mathbf{I_{r\times r}}$ is a $r \times r$ identity matrix; $\mathbf{0_{r\times s-r}}$ is a $r \times s - r$ matrix of zeros; and $\mathbf{e_t}$ is a sequence of uncorrelated (mean zero, covariance Σ) r-dimensional random vectors, the innovation process. The first of the two equations is known as the state-transition equation and the second the observation equation.

An outline of the state space modelling procedure is as follows. A sequence of vector autoregressive models of increasing order is fitted to the data using Yule–Walker equations. Akaike's information criterion is calculated for every model and the order of the model with minimum AIC identified. This determines the number of lags into the past to be used in a canonical correlation analysis. The sample canonical contribution of the past with variables an increasing number of steps into the future is considered; an AIC-type criterion with penalty weight λ weighs the contribution of the canonical correlation of a variable (if $\lambda = 2$ the criterion is the AIC criterion), and variables with a large contribution are added to the state space vector. Once the form of the state space vector is determined, the free parameters are estimated by maximum likelihood based on the sample covariance matrix (or at least an approximation to it). Alternatively, conditional least squares estimates can be obtained.

As the penalty function decreases, the size of the state space vector is non-decreasing. It is possible to develop models for a range of values of the penalty parameter and select a forecasting model on the basis of forecasting performance on a hold-out sample. Usually this boils down to trading goodness-of-fit for improved forecasting performance. Estimation is carried out by both available methods (ML and CLS), in order to investigate fit and forecasting performance for the two methods. Once parameter estimates and estimates of their standard errors are obtained, a backward elimination of parameters that do not contribute significantly to the model is performed. Parameters are eliminated one at a time in order of their t-ratio significance levels if the relative change in their AIC indicates non-significance. Models are re-estimated after the deletion of each parameter.

Different state space models are referred to by the associated penalty parameter, e.g. ST(1.5) refers to the state space model with $\lambda = 1.5$. The results relating to the fit of the various models are presented in Table 5. The forecasting results appear in Tables 6 and 7.

The following features should be noted with regard to model fit:

- CLS performs better than ML estimation, particularly in multivariate terms.
- Contrary to intuition, ML estimation results do not follow a pattern of improved fit with increasing number of parameters. This is an artefact of the optimisation algorithm given the nature of the data.
- Using CLS, fit improves as the state space dimension increases, as expected.

It is clear from these tables, that as far as this study is concerned, CLS provides more reliable estimates than ML and does not suffer from the same instability problems. It should be noted that the improvement in generalised variance and generalised mean square residual is approximately in line with the improvement in product of component variances (respectively, mean squared residual terms). This suggests that the determinants are proportional to the

Table 5
Measures of fit for the multivariate state space model (1980–92)

Model	Estimation method	Univariate measure (Residual standard deviation)			
		France	Germany	Italy	UK
ST(5)	ml	0.28	0.33	0.46	0.20
ST(2.3)	ml	0.24	0.33	0.46	0.20
ST(2.2)	ml	0.23	0.29	0.51	0.21
ST(2.0)	ml	0.23	0.28	0.46	0.21
ST(1.9)	ml	0.25	0.29	0.51	0.20
ST(1.8)	ml	0.24	0.34	0.47	0.20
ST(1.75)	ml	0.46	0.78	0.56	0.22
ST(1.5)	ml	0.31	0.31	0.46	0.21
ST(5)	cls	0.28	0.33	0.46	0.20
ST(2.3)	cls	0.24	0.33	0.46	0.20
ST(2.2)	cls	0.25	0.27	0.47	0.20
ST(2.0)	cls	0.23	0.27	0.46	0.20
ST(1.9)	cls	0.23	0.27	0.46	0.20
ST(1.8)	cls	0.23	0.28	0.46	0.19
ST(1.75)	cls	0.21	0.27	0.46	0.20
ST(1.5)	cls	0.22	0.24	0.44	0.16
		Multivariate measures			
		Generalised MSR $\times 10^{-5}$	Total MSR	Generalised var $\times 10^{-5}$	Total variance
ST(5)	ml	6.35	0.44	6.35	0.44
ST(2.3)	ml	4.60	0.42	4.60	0.42
ST(2.2)	ml	4.78	0.44	4.78	0.44
ST(2.0)	ml	3.44	0.38	3.44	0.38
ST(1.9)	ml	4.46	0.44	4.46	0.44
ST(1.8)	ml	5.01	0.43	5.01	0.43
ST(1.75)	ml	9.95	0.53	9.82	0.53
ST(1.5)	ml	8.04	0.45	8.02	0.45
ST(5)	cls	6.35	0.44	6.34	0.44
ST(2.3)	cls	4.60	0.42	4.59	0.42
ST(2.2)	cls	3.11	0.39	3.11	0.39
ST(2.0)	cls	3.12	0.38	3.12	0.38
ST(1.9)	cls	2.94	0.38	2.94	0.38
ST(1.8)	cls	2.84	0.38	2.83	0.38
ST(1.75)	cls	2.32	0.37	2.32	0.37
ST(1.5)	cls	1.07	0.32	1.07	0.32

Note: ST(1.75) yielded an unstable model using maximum likelihood if backward elimination was continued until there was no further justification to drop parameters. In this case (but not when using conditional least squares estimation) the process was continued only until the last stable solution was found.

product of the diagonal terms, and that the larger models merely improve the fit along the main diagonal without gaining much from dependence between component series.

The following features regarding the forecasting results should be noted:

- There is no consistent pattern in the forecasting performance of the models as the penalty parameter decreases when the ML algorithm is used.
- The forecasting performance using ML is generally poor when the penalty parameter is equal to 1.75 or 1.5, particularly with respect to the German component of the vector series.
- The best forecasting performance in terms of multivariate criteria when using ML is ascribable to the models ST(1.9) and ST(1.8), the former performing better in terms of mean square predic-

Table 6
Means and standard deviations of forecast error for the multivariate state space model (1993–94)

Model	Estimation method	France	Germany	Italy	UK
ST(5)	ml	−0.49	0.52	0.67	0.23
		0.37	0.29	0.62	0.18
ST(2.3)	ml	−0.33	0.53	0.67	0.50
		0.33	0.29	0.62	0.23
ST(2.2)	ml	−0.21	0.48	1.01	0.66
		0.30	0.29	0.70	0.26
ST(2.0)	ml	−0.38	0.54	0.81	0.63
		0.35	0.28	0.65	0.25
ST(1.9)	ml	−0.21	0.59	0.65	0.48
		0.30	0.29	0.61	0.23
ST(1.8)	ml	−0.29	0.67	0.58	0.49
		0.33	0.29	0.60	0.22
ST(1.75)	ml	−0.34	1.59	0.82	0.53
		0.34	0.50	0.62	0.23
ST(1.5)	ml	−0.25	1.05	0.47	0.54
		0.32	0.41	0.59	0.25
ST(5)	cls	−0.49	0.53	0.64	0.50
		0.37	0.29	0.61	0.23
ST(2.3)	cls	−0.30	0.53	0.65	0.50
		0.32	0.29	0.62	0.23
ST(2.2)	cls	−0.23	0.66	0.66	0.53
		0.31	0.30	0.62	0.23
ST(2.0)	cls	−0.37	0.57	0.76	0.49
		0.34	0.28	0.64	0.23
ST(1.9)	cls	−0.24	0.60	0.87	0.48
		0.31	0.29	0.66	0.23
ST(1.8)	cls	−0.19	0.49	0.87	0.42
		0.30	0.27	0.66	0.22
ST(1.75)	cls	−0.26	0.61	0.87	0.50
		0.33	0.28	0.66	0.23
ST(1.5)	cls	−0.36	0.82	0.59	0.62
		0.35	0.31	0.60	0.22

tion error and the latter with respect to forecast variance. The best forecasting performance in terms of univariate criteria varies according to which component is considered.

- In view of the doubt concerning convergence of the ML estimates, and the uneven fit and forecasting performance of the different models using the ML algorithm, it would be inadvisable to use any of these without further testing on data held back for this type of eventuality.
- A somewhat clearer picture emerges on considering the multivariate criteria forecasting performance when using CLS. The general impression is one of at first improving then deteriorating performance as the penalty parameter decreases. Again, univariate performance is fairly mixed, as is to be expected when using models estimated so as to optimise a multivariate criterion.
- With respect to multivariate criteria when using CLS optimisation, the model ST(2.3) seems to exhibit the best forecasting behaviour, with ST(2.2) a close second. This suggests that the 'best' forecasting models may be obtained with a penalty parameter somewhere between the values 2 and 5 (as suggested in SAS documentation).
- The best forecasting performance in terms of multivariate criteria for models estimated by ML and CLS is very much alike. This and the more

Table 7
Multivariate measures of forecast error for the multivariate state space model (1993–94)

Model	Estimation method	Generalised PE $\times 10^{-4}$	Total PE	Generalised variance	Total var $\times 10^{-5}$
ST(5)	ml	3.81	1.87	4.36	0.66
ST(2.3)	ml	2.51	1.71	3.21	0.63
ST(2.2)	ml	4.07	2.46	4.47	0.73
ST(2.0)	ml	3.32	2.16	3.53	0.68
ST(1.9)	ml	2.30	1.66	2.92	0.61
ST(1,8)	ml	2.52	1.71	2.71	0.60
ST(1.75)	ml	10.36	4.40	5.86	0.81
ST(1.5)	ml	6.53	2.36	4.99	0.68
ST(5)	cls	3.83	1.83	4.36	0.65
ST(2.3)	cls	2.37	1.66	3.06	0.62
ST(2.2)	cls	2.41	1.83	2.66	0.63
ST(2.0)	cls	2.62	1.94	3.01	0.65
ST(1.9)	cls	2.48	2.07	2.94	0.67
ST(1.8)	cls	2.41	1.87	3.19	0.66
ST(1.75)	cls	3.34	2.11	2.41	0.67
ST(1.5)	cls	4.48	2.16	2.89	0.62

consistent pattern of fit and forecasting performance with respect to penalty parameter changes suggest that attention may reliably be restricted to CLS as far as this study is concerned. The model ST(2.3) using CLS represents a reasonable compromise between fit and forecasting performance and may be regarded as 'best' as far as multivariate state space forecasting is concerned (in the current study).

5.2. Diagnostic checking

Investigation of the observed residuals from the above model reveals some evidence of non-normality. Moreover an investigation of cross-correlations between the residual series leads to the conclusion that the fitted multivariate state space model has not resulted in a model with an uncorrelated observed innovation process. To summarise, the observed residuals after fitting this multivariate state space model cannot be considered a realisation from a multivariate white noise process and the model is not adequate in this sense. If additional parameters are introduced to remove this inadequacy it is at the cost of forecast performance and may even be at the cost of fit performance. The problem is clearly too many rather than too few parameters. The forecasts for the German and UK series consistently underestimate the observed series. This is on the whole also the case for the Italian series, whereas the French series is generally overestimated (apart from the initial forecasts). The pattern of forecasts shows fairly good correspondence for all series apart from the German series.

6. Conclusions

6.1. Comparison of multivariate and univariate forecasting models

As interest is focused on the development of forecasting models, and to keep discussion to a minimum, attention in what follows is restricted to what may be regarded as the best models according to each approach, i.e. a comparison is made of the results for: the multivariate state space model ST(2.3) with CLS estimation; the set of $(0,1,0)\times(1,1,1)_{12}$ univariate ARIMA models; and the set of univariate state space models with CLS estimation, AIC parameter elimination and fixed state space dimension. A summary of the fit performance of these models with respect to various criteria is presented in Table 8. The following features may be noted:

Table 8
Measures of fit and forecasting performance for selected models in each category

	ARIMA	Univariate state space	ST(2.3)
Measure of fit			
Generalised MSR $\times 10^{-5}$	1.27	0.91	4.60
Generalised var $\times 10^{-5}$	1.28	0.90	4.59
Total MSR	0.31	0.29	0.42
Total variance	0.32	0.29	0.42
St deviation France	0.20	0.19	0.24
Sq root MSR France	0.20	0.19	0.24
St deviation Germany	0.26	0.24	0.33
Sq root MSR Germany	0.26	0.24	0.33
St deviation Italy	0.42	0.42	0.46
Sq root MSR Italy	0.42	0.42	0.46
St deviation UK	0.18	0.17	0.20
Sq root MSR UK	0.18	0.17	0.20
Measure of forecast performance			
Generalised PE $\times 10^{-5}$	3.54	10.13	23.73
Generalised FE var $\times 10^{-5}$	0.38	1.30	3.06
Total PE	0.60	0.77	3.06
Total PE variance	0.21	0.30	0.62
Mean PE France	−0.11	−0.30	−0.30
St deviation PE France	0.20	0.33	0.32
Sq root PE France	0.23	0.45	0.44
Mean PE Germany	0.36	0.39	0.53
St deviation PE Germany	0.22	0.25	0.29
Sq root PE Germany	0.42	0.46	0.60
Mean PE Italy	−0.22	0.24	0.65
St deviation PE Italy	0.30	0.30	0.62
Sq root PE Italy	0.38	0.38	0.90
Mean PE UK	0.46	0.43	0.50
St deviation PE UK	0.17	0.19	0.23
Sq root PE UK	0.49	0.47	0.55

- The set of univariate state space models is uniformly best in terms of fit, the set of ARIMA models uniformly second best and the multivariate state space model uniformly worst.
- The decrease in fit performance between second best and third best is greater than the difference between the best and second best.

The forecasting performance of the models is summarised in Tables 8 and 9. Measures of performance previously used in this study appear in Table 8, whereas more commonly used measures appear in Table 9. The following features may be noted:

- Mean forecast error is significant for all models and all component series, i.e. all models produced a set of biased forecasts for all tourist flows.
- The multivariate state space model performed uniformly worst in respect of bias as measured by mean forecast error and the measures in Table 9. A comparison of the results for the other two models is not as clear-cut. On balance ARIMA modelling appears to be best as it performs consistently best in all but the UK series.
- As far as forecast error variance is concerned, multivariate state space models perform uniformly worst. In terms of univariate and multivariate forecast error variance, ARIMA modelling is best.
- Mean sum of squared prediction error, incorporating both bias and forecast error variance, is perhaps the natural measure to use for comparative purposes. Again, multivariate state space modelling performs not only uniformly worst, but substantially worse than the other models.

Table 9
Univariate measures of forecasting performance %

Origin country		Model		
		Multivariate state space	ARIMA	Univariate state space
France	APE	−3.07	−1.41	3.95
	ME	−4.05	−1.47	−4.02
	MAE	4.98	2.57	4.99
	RMSE	5.87	2.92	5.85
Germany	APE	7.15	4.89	5.31
	ME	7.07	4.84	5.26
	MAE	7.07	4.90	5.30
	RMSE	7.99	5.62	6.15
Italy	APE	15.07	−6.83	7.39
	ME	20.74	−7.24	7.43
	MAE	24.00	9.87	8.90
	RMSE	29.28	11.51	11.80
UK	APE	8.89	6.26	5.86
	ME	6.79	6.30	5.91
	MAE	6.79	6.30	5.91
	RMSE	7.54	6.77	6.50

Notes: APE = aggregate prediction error, RMSE = root mean squared error, MAE = mean absolute error, ME = mean error.

In terms of univariate and multivariate prediction error, ARIMA modelling is clearly best. Comparative plots of the forecast and actual series show that the ARIMA forecasts provide good descriptions of the observed French series over the time period, although tending to overestimate marginally. Both state space model sets of forecasts exhibit substantial drift from the observed French series. None of the sets of forecasts provides a good description of the German series (even though they are very similar). Not only do they underestimate the German series, but they fail to capture the pattern of the observed series. There has clearly been an increase in the exponential growth (given the log term) of German tourist numbers, and there may well have been a change in the pattern of visits to the Seychelles. ARIMA and univariate state space model forecasts provide a marginally better description than the multivariate state space model forecasts, but they are far from satisfactory. Forecasts from the two univariate models provide a good description of the observed Italian series, but the forecasts generated by multivariate state space modelling drift away and generally underestimate the series. Remarkably similar sets of forecasts are obtained for all models for the UK series, and they consistently underestimate the actual series. It seems that the exponential growth in UK tourist numbers has been faster over the later years, and there also appears to have been a change in the pattern of visits to the Seychelles.

6.2. ‘Best’ forecasting model

Taking all of the empirical results into account, the moving average ARIMA models seem to be the ‘best’ if a specific choice of forecasting model has to be made. It produces not only forecasting performance as good as or better than the competitor models but also exhibits good fit performance. The simplicity of structure and resultant ease of interpretation are additional commendable features.

6.3. Structure of series and application area

The final choice of forecasting model, and the process that led to the choice, suggest a number of hypotheses as far as the application area, tourism demand, is concerned. The differenced vector of transformed numbers of tourist arrivals can be expressed in terms of a multivariate innovation process, uncorrelated at non-zero lags and at best weakly cross-correlated at lag 0, and with 'intercept' term identically zero. This suggests that if the transformed vector is decomposed according to the usual additive model, the possible explanatory variables need to exhibit the following general features:

- The contribution of the explanatory factors to the random fluctuation term should be of the same type as the working series itself, i.e. the differenced vector of component contributions should be expressible in terms of a multivariate white noise process. The contribution of a common effect to the four components should result in high cross-correlations at lag 0. There is no evidence of this in our example. This implies that any economic factor cannot be strongly correlated both with the number of tourists and across origin country series. If an economic influence is strongly correlated across origins then the contribution to the random fluctuation term must be negligible; and if an economic influence does have a statistically significant impact then it has to be largely uncorrelated across origin country series. This means that the components of the explanatory variables, such as origin country personal disposable income and the price of a holiday to the destination (Witt & Witt, 1995), that are likely to have an impact on random fluctuations need to be independent if they are to be of any significance. It is useful to know that in improving the forecasting model one may ignore economic (and other) variables as potential explanatory factors if they are highly correlated across origins (with respect to the differenced working series).
- Given that the intercept term in the model is zero after two differences, the contribution of the explanatory factors to the remainder of the decompositional model is locally linear in time. The investigation of locally linear trends and to what extent explanatory variables are common to the components of the series falls within the area of long-term forecasting or econometric modelling.

As has been pointed out previously, the structure of the selected forecasting model for the components of the series is remarkably similar. This reinforces the above conclusions.

References

Akaike, H. (1976). Canonical correlations analysis of time series and the use of an information criterion. In Mehra, R., & Lainiotis, D. G. (Eds.), *Advances and case studies in system identification*. New York: Academic Press, pp. 27–96.

Allen, P. G., & Fildes, R. (2001). Econometric forecasting. In Armstrong, J. S. (Ed.), *Principles of forecasting: a handbook for researchers and practitioners*. Boston: Kluwer Academic Publishers, pp. 303–362.

Aoki, M. (1987). *The state space modeling of time series*. New York: Springer-Verlag.

Bidarkota, P. V. (1998). The comparative forecast performance of univariate and multivariate models: an application to real interest rate forecasting. *International Journal of Forecasting*, *14*, 457–468.

Box, G. E. P., & Jenkins, G. (1976). *Time series analysis: forecasting and control*. San Francisco: Holden-Day.

Brodie, R. J., & de Kluyver, C. A. (1987). A comparison of the short term forecasting accuracy of econometric and naïve extrapolation models of market share. *International Journal of Forecasting*, *3*, 423–437.

Garcia-Ferrer, A., & Queralt, R. A. (1997). A note on forecasting international tourism demand in Spain. *International Journal of Forecasting*, *13*, 539–549.

Garcia-Ferrer, A., Del Hoyo, J., & Martin-Arroyo, A. S. (1997). Univariate forecasting comparisons: the case of the Spanish automobile industry. *Journal of Forecasting*, *16*, 1–17.

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

Granger, C. W. J., & Newbold, P. (1977). *Forecasting economic time series*. San Diego: Academic Press.

Kim, S., & Song, H. (1998). Analysis of inbound tourism demand in South Korea: a cointegration and error correction approach. *Tourism Analysis*, *3*, 25–41.

Kulendran, N., & King, M. L. (1997). Forecasting international quarterly tourist flows using error-correction and time-series models. *International Journal of Forecasting*, *13*, 319–327.

Kulendran, N., & Witt, S. F. (2001). Cointegration versus least squares regression. *Annals of Tourism Research*, *28*, 291–311.

Song, H., & Witt, S. F. (2000). *Tourism demand modelling and forecasting: modern econometric approaches*. Oxford: Pergamon.

Stroud, T. W. F., Sykes, A. M., & Witt, S. F. (1998). Forecasting a collection of binomial proportions in the presence of covariates. *International Journal of Forecasting*, *14*, 5–15.

Turner, L. W., Kulendran, N., & Fernando, H. (1997). Univariate modelling using periodic and non-periodic analysis: inbound tourism to Japan, Australia and New Zealand compared. *Tourism Economics*, *3*, 39–56.

Turner, L. W., Kulendran, N., & Pergat, V. (1995). Forecasting New Zealand tourism demand with disaggregated data. *Tourism Economics*, *1*, 51–69.

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: a review of empirical research. *International Journal of Forecasting*, *11*, 447–475.

Biographies: Johann DU PREEZ is a Senior Consultant with the Advanced Analytical Consultancy Group, ACNielsen, Oxford. His research interests are applied probability, statistical inference and time series analysis.

Stephen F. WITT holds the Chair of Tourism Forecasting in the School of Management, University of Surrey, and is Senior Research Fellow at the Centre for Regional and Tourism Research, Denmark. He has also recently held visiting research positions in Sweden and Australia. His major research interests are econometric modelling of international tourism demand, and assessment of the accuracy of different forecasting methods within the tourism context. He has published over 120 journal articles and book chapters, as well as 13 books including: *Modeling and Forecasting Demand in Tourism* (1992), *Tourism Demand Modelling and Forecasting: Modern Econometric Approaches* (2000), *Tourism Forecasts for Europe 2001–2005* (2001), and *Pacific Asia Tourism Forecasts* 2002–2004 (2002).