Available online at www.sciencedirect.com
ScienceDirect

International Journal of Forecasting 27 (2011) 902–922

international journal of forecasting

www.elsevier.com/locate/ijforecast

Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures

Robert Fildes[a,*], Yingqi Wei[b], Suzilah Ismail[c]

[a] Department of Management Science, Lancaster University, Lancaster, LA1 4YX, UK
[b] The York Management School, University of York, York, YO10 5DD, UK
[c] Faculty of Information Technology and Quantitative Science, Institut Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia

Available online 13 August 2009

Abstract

Airline traffic forecasting is important to airlines and regulatory authorities. This paper examines a number of approaches to forecasting short- to medium-term air traffic flows. It contributes as a rare replication, testing a variety of alternative modelling approaches. The econometric models employed include autoregressive distributed lag (ADL) models, time-varying parameter (TVP) models and an automatic method for econometric model specification. A vector autoregressive (VAR) model and various univariate alternatives are also included to deliver unconditional forecast comparisons. Various approaches for taking into account interactions between contemporaneous air traffic flows are examined, including pooled ADL models and the enhanced models with the addition of a "world trade" variable. Based on the analysis of a number of forecasting error measures, it is concluded that pooled ADL models that include the "world trade" variable outperform the alternatives, and in particular univariate methods; and, second, that automatic modelling procedures are enhanced through judgmental intervention. In contrast to earlier results, the TVP models do not improve accuracy. Depending on the preferred error measure, the difference in accuracy may be substantial.

Keywords: Airline traffic; Comparative forecasting accuracy; Econometric model building; Time-varying parameter; Pooled cross-section time series; Replication

1. Introduction and hypotheses

The Air Transport industry is an increasingly important component of national and global economies. Its development involves enormous capital investments on the part of all of the agents concerned. Because of the perishable nature of the product (once the aircraft takes off, the empty seats are considered as an opportunity cost to the airline), forecasting air traffic flows in both the short and medium terms is an important factor in increasing the profitability of airlines, and a critical part of transport planning by air transport authorities and government bodies.

* Corresponding author. Tel.: +44 1524 593879.
E-mail addresses: r.fildes@lancaster.ac.uk (R. Fildes), yw542@york.ac.uk (Y. Wei).

0169-2070/$ - see front matter
doi:10.1016/j.ijforecast.2009.06.002

A number of papers have attempted to develop models suitable for forecasting air traffic flows, including Abed, Ba-Fail, and Jasimuddin (2001), Anderson and Kraus (1980), Fridström and Thune-Larsen (1989), Grubb and Mason (2001), Ippolito (1981), Jorge-Calderon (1997), Kaemmerle (1991), Matsumoto (2004) and Young (1972). More recently, Blunk, Clark, and McGibany (2006) and Lai and Lu (2005) have examined the effects of the September 11, 2001, terrorist incident using demand models. Because of the importance of infrastructure planning, various agencies with planning and regulatory responsibilities for air traffic infrastructure, as well as private companies such as aircraft manufacturers and airlines, have also developed their own models. These models are largely based on the approach proposed by Quandt and Baumol (1966), who showed that socio-economic variables are important in modelling point-to-point air traffic flows between countries.

However, these earlier studies have provided only limited evidence on forecasting accuracy and the comparative performance of alternative model specifications. In particular, many of these forecasting studies have contented themselves with univariate methods, and many earlier evaluations have been limited to only forecast horizons which are too short to be valuable for planning purposes. This problem is exacerbated by looking at the comparative performance, which may be dependent on the forecast horizon.

In addition, recent years have seen the development of a number of alternative econometric modelling approaches that offer the prospect of enhanced forecasting accuracy, but have rarely been compared (Allen & Fildes, 2001). These approaches include Hendry's general-to-specific approach to model building with a full lag structure (Hendry, 1986), a model that includes a full lag structure and pools cross-sectional evidence, a time-varying parameter (TVP) model (Garcia-Ferrer, Highfield, Palm, & Zellner, 1987), and a PcGive automatic selection model based on Hendry's methodology (Hendry & Krolzig, 2001). Replicating and extending the limited studies published so far is important for generalising about the conditions under which the results hold (Hubbard & Vetter, 1996). In light of the importance of air traffic forecasting, the paucity of evidence on the relative benefits of these alternative approaches, and the lack of insight into what leads to differences in relative performance, this paper compares the accuracy of the above four approaches using price, income and trade as potential explanatory variables. An enhanced version of these models incorporating a "world trade" variable to capture the overall growth in demand across all countries is also included. These are contrasted with 'naïve' univariate alternatives, as well as with a vector autoregressive (VAR) model. Our aim is to offer further evidence on the conditions under which econometric methods outperform time series alternatives. Recent aggregate evidence from Athanasopoulos, Hyndman, Song, and Wu (2008) has shown no overall improvement from using econometric methods with tourism data series, apparently similar to those considered here.

The choice of econometric models to include in the comparisons has been based on their strong performances in earlier empirical studies (for example, Garcia-Ferrer et al., 1987). In general, for non-financial series, earlier research has shown that such econometric models outperform the autoregressive and naïve benchmarks (Allen & Fildes, 2001), but the evidence is not overwhelming, and our principal aim is to provide further reliable evidence.

We also wish to investigate a number of subsidiary hypotheses.

(i) Model specification is usually overly subjective (Pagan, 1999), but does the subjectivity of the expert modeller compared to the automatic modelling approach embodied in PcGive lead to improvements in accuracy? Whilst there is substantial evidence that econometric model-based forecasts using the same information set outperform judgement (Dawes, Faust, & Meehl, 1989), to the best of our knowledge, there is no study that directly addresses this question of objective versus subjective specifications, with the closest studies being those concerned with the specification of ARIMA models. However, *a priori*, the same limitations of judgemental forecasts would apply to judgemental model building, and the automatic approach should lead to improvements.

(ii) The consideration of interactions between contemporaneous air traffic flows, through both the inclusion of a "world trade" variable and estimation using a Seemingly Unrelated Regression (SUR) approach, offers the prospect of improved accuracy, but, as Du Preez and Witt (2003) show, this is far from inevitable. Following on from Zellner, Hong, and Min's (1991) conclusion as to the benefits of including a generic variable that acts as a proxy for the many unobserved explanatory variables, we hypothesize that both of these approaches should lead to accuracy improvements.

(iii) Although a relatively neglected approach to econometric model specification, the inclusion of time-varying parameters has typically led to improved forecasting accuracy (Allen & Fildes, 2001; Garcia-Ferrer et al., 1987). A recent example in tourism forecasting (Li, Song, & Witt, 2006; Li, Wong, Song, & Witt, 2006) again demonstrated improved performance, including at longer forecast horizons. In the airline industry, because of its changing structure over the 40 year history we study, *a priori* we would expect the same conclusions to hold, and this will also be examined.

(iv) Previous research has emphasized the potential importance of the choice of error measure (Fildes & Ord, 2002). We will also provide further evidence on this issue.

The plan of the paper is as follows. Section 2 describes the variables and data. Section 3 presents the models, methodology and error measures. The results based on ex-post forecasting up to three years ahead are discussed in Section 4. Finally, Section 5 draws conclusions on the relative forecasting accuracy of the particular methods.

2. Variables and data

The dependent variable is the growth rate of the demand for air travel. It is measured by the percentage change (i.e., in logged differences) in the total annual international passenger traffic (000s) between country pairs that always include UK airports as either the origin or the destination. The other countries are Germany, Sweden, Italy, USA and Canada. These countries are chosen to avoid the routes of highest leisure tourist intensity, where the determinants are more subject to fluctuating changes in consumer tastes, such as Spain, the country that attracted the largest numbers of travellers over the period under study (National Statistics Office, 2002). Canada, USA and Italy attract higher proportions of leisure tourists than Germany and Sweden. Data are obtained from the Annual Statement of Movements (Passenger and Cargo) published by the UK's Civil Aviation Authority (CAA) for 1961–2002. They include all passengers carried on scheduled and chartered services by all airlines flying between the UK and each of the five countries, excluding those carried on chartered services by government departments. The figures do not make any distinction between economy and business classes.

The process of selecting the explanatory variables could start with several potential variables, such as the growth rates of gross domestic product (GDP), population, employment/unemployment rates, airfares, and volume of trade. To maintain the simplicity of the model formulation, the number of explanatory variables is narrowed down to include only the growth rates of income, trade and price, which have been proved to be important in earlier studies of the demand for air travel, for example, Kaemmerle (1991) and Quandt and Baumol (1966). When income increases, more people can afford to fly. When airfares increase, the demand for air travel is likely to lessen. Trade is included to capture the effects of passengers who travel for business purposes. Swan (2002) notes that travel grows with trade. The explanatory variables are aggregated to provide a single route-based measure, using a weighted average in line with passenger flows. Annual data are used, as more frequent data on passenger movements and price are not available for the whole period, and our focus is on longer term forecasting. Details about data sources and the data transformation and weighting procedure are given in the Appendix.

Fig. 1 shows the nature of the series pertaining to the growth of demand for air travel between the UK and the five countries, as well as world trade.

Prior to estimation, the data were tested for stationarity using the Augmented Dickey–Fuller (ADF) test. The results of the test show that all variables under investigation are stationary, that is, I(0).[1] Descriptive statistics and the correlation matrix are presented in Table 1.

[1] The model is specified in growth rate variables, which are stationary (I(0)). The first-difference model may be over-restrictive compared with a model with a cointegrating relationship or a model in levels. However, both empirical and theoretical work shows that a differenced model will produce more robust forecasts. The specification also permits a direct comparison with the work of Garcia-Ferrer et al. (1987).

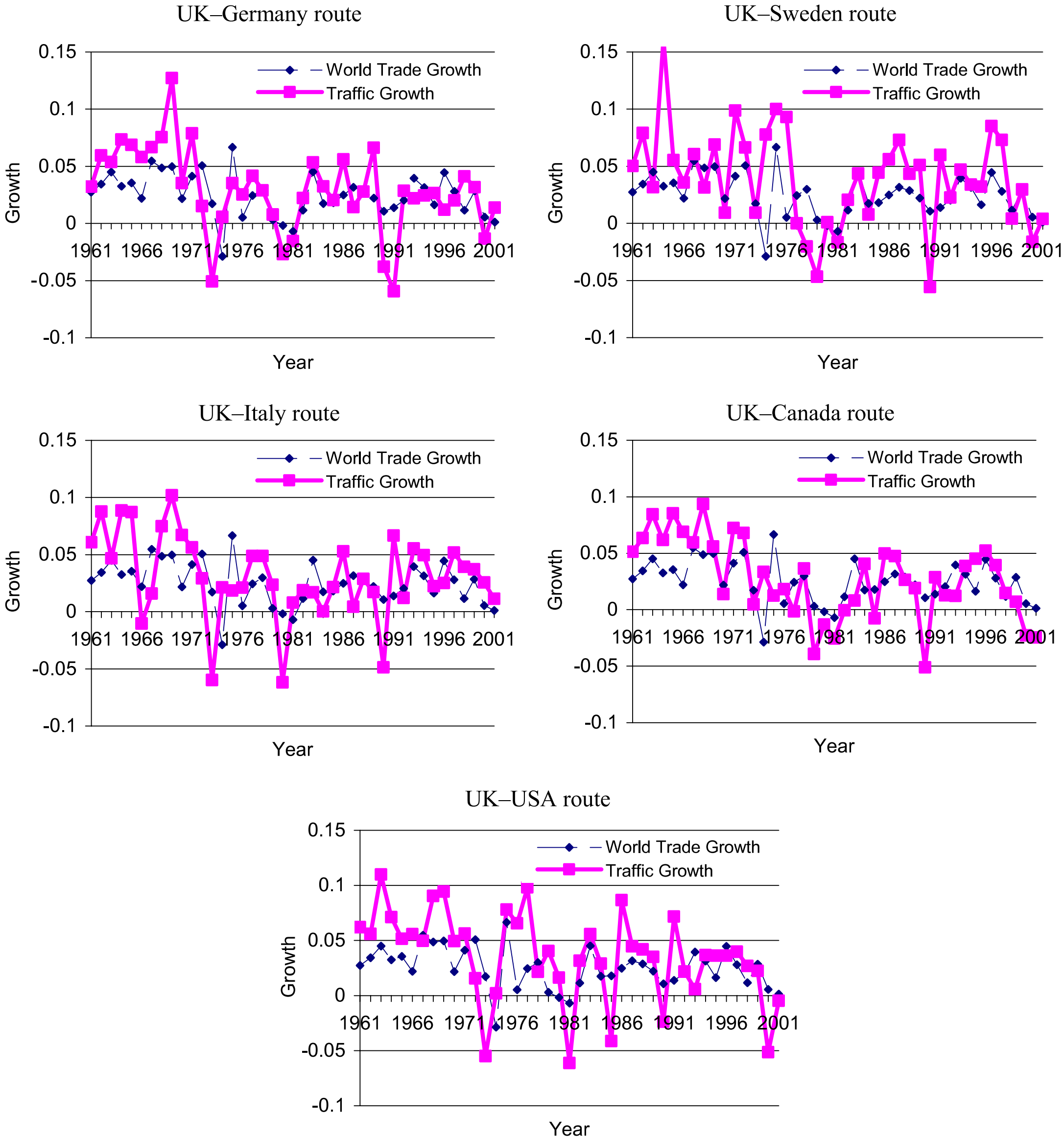

Fig. 1. Growth rates of annual international passenger traffic by country and world trade (the solid line represents traffic growth, and the dashed line world trade).

3. Methodology

Various econometric methodologies have been proposed in the literature as being suitable for developing effective forecasting models. Here we consider those that have empirically proved the most effective, namely the general-to-specific approach associated with the London School of Economics (Clements & Hendry, 1998; Pagan, 1987), Zellner's Structural Equation Modelling Time Series Analysis approach (SEMTSA) (Zellner & Palm, 2004), and time-varying parameters (TVP) models. A VAR model is also included to provide unconditional forecasts. Benchmark autoregressive models and exponential smoothing are included, along with two naïve models, in order to measure the improvements arising from the use of more complex models and additional explanatory variables.

Table 1
Descriptive statistics and correlation matrix.

	Mean	Median	Max.	Min.	s.d.	Correlation			
						Δy_t	Δx_{1t}	Δx_{2t}	Δx_{3t}
UK–Canada									
Δy_t	0.0640	0.0658	0.2153	−0.1175	0.0808	1			
Δx_{1t} (Income)	0.0111	0.0206	0.2253	−0.1638	0.0739	0.29	1		
Δx_{2t} (Trade)	−0.0217	−0.0172	0.1733	−0.4185	0.1200	0.41	0.57	1	
Δx_{3t} (Ticket price)	−0.0262	−0.0182	0.4671	−0.4145	0.1763	−0.34	0.33	−0.07	1
Δx_{4t} (World trade)	0.0575	0.0572	0.1532	−0.0661	0.0430	0.57	0.14	0.32	−0.07
UK–Germany									
Δy_t	0.0673	0.0655	0.2930	−0.1369	0.0834	1			
Δx_{1t} (Income)	0.0296	0.0344	0.2778	−0.1981	0.0984	0.13	1		
Δx_{2t} (Trade)	0.0522	0.0553	0.2958	−0.2110	0.1006	0.52	0.56	1	
Δx_{3t} (Ticket price)	−0.0082	−0.0065	0.2190	−0.2033	0.0818	0.00	0.52	0.31	1
Δx_{4t} (World trade)	0.0575	0.0572	0.1532	−0.0661	0.0430	0.56	0.05	0.40	−0.08
UK–Italy									
Δy_t	0.0720	0.0587	0.2346	−0.1423	0.0837	1			
Δx_{1t} (Income)	0.0056	0.0288	0.2136	−0.2573	0.1018	0.49	1		
Δx_{2t} (Trade)	0.0451	0.0471	0.2812	−0.3112	0.1245	0.47	0.68	1	
Δx_{3t} (Ticket price)	−0.0326	−0.0193	0.2862	−0.8081	0.1901	−0.15	0.26	0.00	1
Δx_{4t} (World trade)	0.0575	0.0572	0.1532	−0.0661	0.0430	0.40	0.14	0.23	−0.14
UK–Sweden									
Δy_t	0.0902	0.0998	0.3720	−0.1283	0.0972	1			
Δx_{1t} (Income)	0.0089	0.0211	0.1867	−0.1762	0.0869	0.12	1		
Δx_{2t} (Trade)	0.0091	0.0275	0.2422	−0.3494	0.1193	0.29	0.38	1	
Δx_{3t} (Ticket price)	−0.0272	−0.0152	0.3310	−0.3950	0.1419	−0.12	0.72	0.21	1
Δx_{4t} (World trade)	0.0575	0.0572	0.1532	−0.0661	0.0430	0.44	0.12	0.37	−0.02
UK–USA									
Δy_t	0.0825	0.0915	0.2532	−0.1410	0.0940	1			
Δx_{1t} (Income)	0.0273	0.0232	0.2151	−0.2354	0.0614	0.23	1		
Δx_{2t} (Trade)	0.0357	0.0425	0.2272	−0.3841	0.1173	0.41	0.41	1	
Δx_{3t} (Ticket price)	−0.0189	−0.0081	0.6047	−0.3890	0.1922	−0.21	−0.03	−0.28	1
Δx_{4t} (World trade)	0.0575	0.0572	0.1532	−0.0661	0.0430	0.53	0.19	0.30	0.13

- Autoregressive Distributed Lag (ADL) model

  The general unrestricted Autoregressive Distributed Lag (ADL) model can be written in the form

$$\Delta y_{it} = \alpha_{i0} + \sum_{j=1}^{J} \alpha_{ij} \Delta y_{i(t-j)} + \sum_{k=1}^{K} \sum_{j=0}^{J_k} \phi_{ikj} \Delta x_{ik(t-j)} + \varepsilon_{it}, \qquad (1)$$

where j is the lag length, $i = 1, 2, \ldots, N$ (countries), $t = 1, 2, \ldots, T$ (time periods), and Δy_{it} is the growth rate of the number of passengers in year t for route i. Δx_{ikt} is the growth rate of the kth explanatory variable in year t for route i. K is the total number of independent variables, which is either 3 or 4, depending on whether or not the "world trade" variable is included. J and J_k are the numbers of lags of the dependent and independent variables, and are set to be 3 in the paper. ε_{it} are identically independently distributed random errors with mean zero and variance σ_ε^2, and α and ϕ are unknown parameter vectors to be estimated.

The general-to-specific methodology for developing a "good" ADL model supposes that 'a general class of possible models' can be specified initially (Charemza & Deadman, 1997). It then proceeds to select the most 'parsimonious model from that class' by the process of elimination, beginning by removing the variable with the highest p-

value, then the variable with the highest p-value in the new estimation, given that the first variable has already been removed, and so on. At each stage, the intercept is included in the estimation and the tests of parameter restrictions are performed, ensuring correct signs of the variables (as a multiplier) according to theory, and the satisfaction of t-tests for individual variables, F-tests for groups of variables, and the test of misspecification. In this study, our starting point is a general autoregressive distributed lag model with 3 lags.

- Pooled ADL model

  An extended version of the ADL model is derived by pooling the data and then estimating all models simultaneously using the seemingly unrelated regression (SUR) approach. This is based on the assumption that there exist common (but unmeasured) influences on air traffic flows which induce contemporaneous correlation among the errors of the individual routes. The SUR approach allows both intercepts and slope coefficients to vary across routes. The SUR model can be represented compactly as

$$\Delta \boldsymbol{Y} = \Delta \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

  where $\Delta \boldsymbol{X}$ is a block diagonal matrix of N cross sections, and each one of $\Delta \boldsymbol{X}_i$ has a dimension $T \times K'$. The K' explanatory variables may include the constant term and the lagged dependent variables. Generalized least squares (GLS) provides an efficient estimator.

- PcGive automatic econometric model selection

  PcGive is a software programme created to overcome some of the subjectivities inherent in model building, which typically relies overmuch on the tacit knowledge of the model builder (Allen & Fildes, 2005; Pagan, 1999). Starting with a ADL model with a 'full' lag structure (known as the general unrestricted model – GUM), the selection process uses the recommended non-stringent significance level (for example, 0.9) to simplify the GUM, whilst ensuring that congruence (where the model passes all of the diagnostic tests) is maintained throughout the process, right up to the final stage where parameter restrictions are rejected and the final model is identified (Hendry & Krolzig, 2001).

  The difference between the ADL and PcGive automatic models lies in the selection process. The former is carried out by the researcher, based on such factors as the correct signs of the coefficients according to the relevant theory and the satisfaction of statistical tests. In contrast, with PcGive automatic modelling, the model is chosen automatically based on a validated model simplification process aimed at achieving a data-congruent encompassing model (Hendry & Krolzig, 2001).

- Time Varying Parameter (TVP) modelling

  In recent years there has been an increased questioning of the assumption that the regression coefficients are constant in models such as those described above. The shift from the assumption of constancy in the effects of the economic drivers to the use of time-varying parameters has occurred because of the recognition of changes in aggregation effects and policies, adaptive optimisation on the part of economic agents (Garcia-Ferrer et al., 1987; Harvey, 1989; Riddington, 1993; Zellner & Min, 1993), and changes in tastes, among other factors. This may prove particularly relevant in an industry such as the airline industry, whose economic role and preferences for travel destinations change over time. Detailed discussions on the TVP model can be found in, for example, Judge, Griffiths, Hill, Lutkepohl, and Lee (1985), but a brief discussion is provided here. The TVP model is usually represented as the observation equation (3) and the state equation (4),

$$\Delta y_{it} = \Delta \boldsymbol{x}'_{it}\boldsymbol{\beta}_{it} + \varepsilon_{it} \tag{3}$$

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{it-1} + \boldsymbol{\nu}_{it}, \tag{4}$$

  where Δy_{it} is the dependent variable and $\Delta \boldsymbol{x}_{it}$ is a vector of explanatory variables. $\boldsymbol{\beta}_{it}$ is a vector of parameters subject to time-dependent variation. ε_{it} and $\boldsymbol{\nu}_{it}$ are identically independently distributed random errors with zero mean and constant variances. The $\boldsymbol{\beta}$s are assumed to be adaptive in nature and subject to permanent and transitory changes, and are modelled in Eq. (4) as a multivariate random-walk without drift. Thus, the latest estimate of $\boldsymbol{\beta}_{it}$ at the forecast origin, time $T_0 - k$, gives greater weight to more recent observations, taking such market changes into account. The forecast of $\boldsymbol{\beta}_{it}$ is the latest estimate, because of the random-walk state equation (4).

- Vector autoregressive (VAR) model

  In a vector autoregressive (VAR) model, endogenous and exogenous variables are not distinguished. Each variable is treated as endogenous and is regressed on its own lagged variables and all other variables in the system. The VAR model can expressed as

$$\begin{aligned}
\Delta y_{it} &= \alpha_{i0} + \sum_{j=1}^{J} \alpha_{ij} \Delta y_{i(t-j)} \\
&\quad + \sum_{k=1}^{K} \sum_{j=1}^{J_k} \phi_{ikj} \Delta x_{ik(t-j)} + \varepsilon_{it} \\
\Delta x_{1it} &= \beta_{i0} + \sum_{j=1}^{J} \beta_{ij} \Delta y_{i(t-j)} \\
&\quad + \sum_{k=1}^{K} \sum_{j=1}^{J_k} \vartheta_{ikj} \Delta x_{ik(t-j)} + \delta_{it} \\
&\vdots \\
\Delta x_{Kit} &= \gamma_{i0} + \sum_{j=1}^{J} \gamma_{ij} \Delta y_{i(t-j)} \\
&\quad + \sum_{k=1}^{K} \sum_{j=1}^{J_k} \xi_{ikj} \Delta x_{ik(t-j)} + \eta_{it}
\end{aligned}$$

  where α, β, γ, ϕ, θ, and ξ are vectors of coefficients. ε_{it},δ_{it}, and η_{it}, are identically independently distributed random errors. As with the ADL model, the maximum lag length considered is 3. All variables have been included in the VAR system.

- Naïve models

  Using a modelling technique similar to that used by Garcia-Ferrer et al. (1987), two benchmark models are used as the basis for a comparison against all other estimated models. These have typically performed surprisingly well, at least in macroeconomic forecasting comparisons (Allen & Fildes, 2001). The two benchmark models are naïve model one (NM1), which assumes zero growth, and naïve model two (NM2), which assumes that the future growth rate equals the rate in the last period. They are defined respectively as

$$NM1 : \Delta \hat{y}_t = 0 \tag{5}$$

$$NM2 : \Delta \hat{y}_t = \Delta y_{t-1}. \tag{6}$$

- Autoregressive of order 3 (AR(3)) model

  Garcia-Ferrer et al. (1987) argue that the poor performance of the NM1 and NM2 models in their study could be improved by fitting an autoregressive model of order 3 (AR(3)). This model is also included here, and is defined as

$$\begin{aligned}
\Delta y_{it} &= \beta_{0i} + \beta_{1i} \Delta y_{i(t-1)} + \beta_{2i} \Delta y_{i(t-2)} \\
&\quad + \beta_{3i} \Delta y_{i(t-3)} + \varepsilon_{it}.
\end{aligned} \tag{7}$$

- Exponential smoothing model

  Exponential smoothing models have regularly performed well in comparative performance evaluations (Fildes & Ord, 2002). Here we have adopted the expert system used in ForecastPro (Goodrich, 2000) for producing the forecasts.

3.1. Operational issues

The forecasting models can be evaluated either conditionally, assuming a knowledge of the exogenous variables, or unconditionally, with the exogenous variables predicted. The latter evaluation provides evidence about the appropriate choice of model in the practical forecasting problem faced by users in the industry. Conditional forecasts are useful for evaluating the relative strengths of the alternative model specifications under consideration. However, such conditional comparisons typically benefit the causal models at the expense of the univariate autoregressive alternatives, although uncertainty in the explanatory variables is typically not the main cause of forecast error (Clements & Hendry, 1998). In addition, therefore, the naive and univariate autoregressive models and the VAR model were included to provide unconditional forecasts.

The models have been estimated initially on information from 1961–1991 ($n = 31$), from which one- to three-year-ahead conditional forecasts were calculated. Rolling forecasts were produced by extending the data base year by year through the hold-out sample, 1992–2002 ($n = 11$), and re-estimating the model on the extended data set. Where the forecasts are conditional, actual xs and actual lagged ys were used when these would be known, and the estimated lagged ys were used for forecast horizons greater than one-period-ahead.

As one important objective of this paper is replication, the precise details of how the models described above have been implemented is included below.

- ADL model and pooled ADL model: 3 period lags of all variables were initially considered for inclusion in an ADL model. These were simplified according to the procedure described above using all of the data. This provides an evaluation of the model that would ideally have been selected, based on a foreknowledge of the data. This has the advantage of providing evidence on the potential accuracy of the ADL model versus the accuracy that was achieved using the automatic procedure. (An alternative approach could have been to apply the operational rules described above to potentially re-specify the model as the available data base is rolled forward.)

  The models' parameters were re-estimated for each rolling origin. A dummy variable was included in the model for Germany, to take into account re-unification in 1991, but it was shown to be insignificant during the process of reducing the parameterisation of a model. For the Pooled ADL model, once the ADL model for each route was determined, they were pooled and estimated using the SUR approach. Eviews 6.0 was used in the estimation.
- PcGive automatic econometric model selection: 3 period lags of all variables were initially considered for inclusion. Simplification and model selection followed the procedures laid down in PcGive™12, and no subjective adjustments were made to take into account apparent model defects. Given the limited amount of data initially available, the model selected was often more complex than when a more extended data base is used. There was also evidence of multicollinearity.
- TVP: For the TVP model estimations, the simplified ADL model structure derived using the general-to-specific methodology is adopted for the observation equation. The coefficients, including that for the constant, are assumed to follow a random walk without drift. The Kalman filter algorithm in Eviews 6.0 was employed. To implement the Kalman filter, we need to know the initial conditions. By default, Eviews 6.0 can handle the initial condition by treating the initial values as diffuse (i.e., the initial one-step-ahead mean = zero; the initial one-step-ahead variance = $10^6 I_M$, following Koopman, Shephard, & Doornik, 1999).
- VAR model: 3 period lags of all variables were initially considered for inclusion in a VAR model. As demonstrated by Lütkepohl (1993), the determination of the lag length of the VAR is a critical element in the specification of VAR models. Overfitting (i.e., selecting a higher order lag length than the true lag length) causes an increase in the mean square forecast errors, while underfitting generates autocorrelated errors. Hence, we select the lag length using the Akaike Information Criterion (AIC), because it is more suitable for short time series than other criteria.
- Exponential Smoothing and ARIMA: These models were run automatically using ForecastPro XE (v.5). An expert system chooses among alternative exponential smoothing specifications, including simple smoothing and Holt's linear trend (Goodrich, 2000). This permits new specifications to be used as the data base changes. For most of the time periods and countries, the chosen model was simple smoothing.

3.2. Error measures

In evaluating the models, no single error measure captures the distributional features of the errors when summarised across data series. A recent summary of some of the issues is given by Hyndman and Koehler (2006). Here, four error measures have been used, which should capture the key characteristics of the results: (i) Root Mean Square (Percentage) Error (RMSE), (ii) Geometric Root Mean Square Error (GRMSE), (iii) Mean Absolute Scaled Error (MASE), and (iv) Geometric Mean Relative Absolute Error (GRelAE).

(i) The RMSEs for l-step-ahead forecasts are generated over n out-of-sample data points ($T_0 + 1$, $T_0 + 2, \ldots, T_0 + n$), and are written as

$$\text{RMSE} = \sqrt{\frac{\sum_{t=T_0}^{T_0+n-l} e_t^2(l)}{n+1-l}}, \tag{8}$$

where $e_t(l) = y_{t+l} - \hat{y}_t(l)$, y_{t+l} is the actual observation at time $t + l$, and $\hat{y}_t(l)$ is the l-step-ahead forecast of y_{t+l} generated from the models using T_0 to $T_0 + n - l$ observations.

(ii) The RMSE can be greatly affected by outliers (Fildes, 1992). As a means of overcoming such a problem, when confronting a large error term due to a particularly bad forecast, Fildes (1992) suggests using the GRMSE.

$$\text{GRMSE} = \left[\prod_{t=T_0}^{T_0+n-l} e_t^2(l)\right]^{\frac{1}{2(n+1-l)}}. \tag{9}$$

Of the two error measures, the RMSE is commonly favoured among academics, despite its lack of robustness in small samples.

(iii) The MASE is a relative error measure, defined as the Mean Absolute Error scaled by the 'naïve' random walk error as follows:

$$\text{MASE} = \frac{\frac{1}{n+1-l}\sum_{t=T_0}^{T_0+n-l} |e_t(l)|}{\frac{1}{T_0-1}\sum_{l}^{T_0} |Y_t - Y_{t-1}|}. \tag{10}$$

It has recently been suggested by Hyndman and Koehler (2006) as a means of overcoming some of the deficiencies of other error measures, such as observations and errors around zero, which may affect our fourth measure, GRelAE.

(iv) Proposed by Fildes (1992) as a means of overcoming problems with outliers, the Geometric Mean Relative Absolute Error (or its equivalent, the Geometric Mean Root Mean Squared Relative Error) is defined as:

$$\text{GRelAE} = \frac{\text{GRMSE}}{\text{GRMSE(Naive)}}, \tag{11}$$

i.e., each error, $|e_t(l)|$, is scaled by the corresponding naïve error, $|Y_t - Y_{t-1}|$, and the geometric mean calculated.

The error measures have been summarised across countries by taking the median. In addition, the geometric mean has been calculated for the two measures, RMSE and GRelAE, which are themselves based on a geometric mean. The rankings stay much the same, and are not reported.

4. Empirical results and forecasting performance

The first 31 observations, 1961–1991, are used to fit the initial models, while the remaining 11 observations, 1992–2002, are used to re-estimate and evaluate the forecasting performance of the models, as we described above. After appropriate simplification, the estimated ADL models for each route (over the period 1961–2002) are shown in Table 2. All of the variables show the expected signs (that is, a positive sign for income and trade,[2] and a negative sign for ticket price). The F tests indicate an overall significance of the coefficients in their respective models. The values of $\overline{R}^2$ range from 0.152 for the UK–Sweden route to 0.469 for the UK–Germany route. All models pass the tests of serial correlation, functional form and heteroscedasticity. The results of Chow tests of parameter constancy in Table 2 are also shown to be statistically insignificant.

Only the UK–Germany, UK–USA and UK–Canada models include lagged dependent variables, which are significant at the 10% level. The income variable is included for the UK–Italy only, with an elasticity less than one. For all other routes, income does not appear to directly influence the demand for air travel significantly. This suggests that, with the rapid development of transportation technologies, air travel is becoming less of a luxury in people's lives and business activities. Four out of five routes show that trade is a more reliable predictor than income. More bilateral trade implies a higher level of integration between countries, and, as a result, an increasing need for air travel between countries. This is consistent with Swan's (2002) study on the US. Price is only included in the UK–Germany, UK–Italy and UK–Canada routes, and it appears that air travel is price inelastic over the time and countries modelled. One possible explanation for this is that there are inevitable errors in the measurement of the price variable, as suggested by Swan (2002), who instead used yield data as a measure for the US, an option which is not available here, due to data limitations. (The correlation between the two data sets in the overlapping years was low but positive.) Also, low-cost airlines had not yet had the impact which has been seen more recently.

One path towards improving the performance of the ADL models is to take into consideration the possible

[2] Trade here refers to the bilateral trade between the UK and the other countries.

Table 2
Estimated ADL models, 1961–2002.

	UK–Germany	UK–Sweden	UK–Italy	UK–USA	UK–Canada
Intercept	0.001	0.082***	0.065***	0.056***	0.0260*
	(0.016)	(0.015)	(0.011)	(0.017)	(0.013)
Δy_{t-1}	0.293**	–	–	0.287*	0.509***
	(0.142)			(0.156)	(0.125)
Δy_{t-2}	–	–	–	–	–
Δy_{t-3}	0.258*	–	–	–	–
	(0.129)				
Δx_{1t}	–	–	0.463***	–	–
(Income)			(0.114)		
Δx_{1t-1}	–	–	–	–	–
Δx_{1t-2}	–	–	–	–	–
Δx_{1t-3}	–	–	–	–	–
Δx_{2t}	0.374***	0.274**	–	0.350***	–
(Trade)	(0.117)	(0.122)		(0.119)	
Δx_{2t-1}	–	–	–	−0.314**	–
				(0.123)	
Δx_{2t-2}	–	0.281**	–	–	0.199***
		(0.124)			(0.085)
Δx_{2t-3}	–	–	–	–	–
Δx_{3t}	−0.286**		−0.129**	–	−0.196***
(Ticket price)	(0.134)		(0.061)		(0.058)
Δx_{3t-1}	–	–	–	–	–
Δx_{3t-2}	−0.415***	–	–	–	–
	(0.139)				
Δx_{3t-3}	0.269*	–	–	–	–
	(0.135)				
$\overline{R}^2$	0.469	0.152	0.282	0.258	0.426
Autocorrelation	5.696	0.272	3.103	0.616	0.467
	[0.058]	[0.873]	[0.212]	[0.735]	[0.792]
Functional form	0.107	1.081	1.790	0.009	0.231
	[0.745]	[0.306]	[0.189]	[0.925]	[0.634]
Heteroscedasticity	9.484	0.152	1.128	0.910	4.263
	[0.148]	[0.927]	[0.569]	[0.823]	[0.235]
Parameter constancy	0.349	0.570	0.634	0.863	0.346
Breakpoint = 1992	[0.923]	[0.639]	[0.598]	[0.497]	[0.845]
F	6.446***	4.395***	8.856***	5.515***	10.42***
AIC	−2.551	−1.888	−2.385	−2.079	−2.642

Notes: The values in parentheses are standard errors; those in square brackets are p-values.
* indicates that the variables are significant at the 10% level.
** indicates that the variables are significant at the 5% level.
*** indicates that the variables are significant at the 1% level.
– means that the variable is not included in the estimation.

interdependencies among the routes. These could arise from common economic and/or social factors across the world, such as political upheavals (for example, Iraq's invasion of Kuwait), as well as unobserved common price movements. A solution to this problem is to consider a common variable believed to affect all of the routes. One possible variable is the total trade of all industrial countries, which is called the "world trade" variable here (Garcia-Ferrer et al., 1987, used the 'world' variable to improve their forecasts of GDP growth rates for 9 countries). Another reason to include 'world data' is to capture the effects of through traffic passengers. 'Through traffic' is defined as traffic originating from another country or airport, and pass-

Table 3
Estimated ADL models with the “world trade” variable, 1961–2002.

	UK–Germany	UK–Sweden	UK–Italy	UK–USA	UK–Canada
Intercept	−0.044**	0.029	0.032**	0.007	−0.02
	(0.022)	(0.023)	(0.018)	(0.019)	(0.014)
Δy_{t-1}	–	–	–	–	–
Δy_{t-2}	–	–	–	–	0.316***
					(0.11)
Δy_{t-3}	0.281**	–	–	–	–
	(0.12)				
Δx_{1t}	–	–	0.363***	–	0.401***
(Income)			(0.108)		(0.125)
Δx_{1t-1}	–	–	–	–	–
Δx_{1t-2}	–	–	–	–	–
Δx_{1t-3}	–	–	–	–	0.221*
					(0.134)
Δx_{2t}	0.275**	–	–	0.223**	–
(Trade)	(0.109)			(0.109)	
Δx_{2t-1}	–	–		–	–
Δx_{2t-2}	–	0.241**	–	–	–
		(0.117)			
Δx_{2t-3}	–	–	–	–	–
Δx_{3t}	−0.29**				−0.218***
(Ticket price)	(0.135)				(0.05)
Δx_{3t-1}	–	–	–	–	–
Δx_{3t-2}	−0.439***			−0.12*	–
	(0.13)			(0.063)	
Δx_{3t-3}	–	–	–	−0.160***	–
				(0.061)	
Δx_{4t}	0.697***	0.974***	0.662**	0.991***	0.781***
(World trade)	(0.245)	(0.319)	(0.256)	(0.281)	(0.188)
Δx_{4t-1}	0.455*	–	–	–	–
	(0.239)				
Δx_{4t-2}	–	–	–	–	–
Δx_{4t-3}	–	–	–	–	–
$\overline{R}^2$	0.535	0.232	0.318	0.395	0.621
Autocorrelation	0.547	0.216	1.775	1.595	0.217
	[0.761]	[0.898]	[0.412]	[0.450]	[0.897]
Functional form	0.205	0.346	0.616	0.297	0.240
	[0.654]	[0.560]	[0.438]	[0.590]	[0.628]
Heteroscedasticity	8.217	2.144	0.113	2.068	4.927
	[0.223]	[0.342]	[0.945]	[0.723]	[0.425]
Parameter constancy	0.193	0.350	0.559	0.083	0.785
Breakpoint = 1992	[0.984]	[0.790]	[0.646]	[0.994]	[0.590]
F	8.085***	6.727***	10.319***	7.050***	13.149***
AIC	−2.683	−1.987	−2.436	−2.304	−3.06

Notes: The values in parentheses are standard errors; those in square brackets are p-values.
* indicates that the variables are significant at the 10% level.
** indicates that the variables are significant at the 5% level.
*** indicates that the variables are significant at the 1% level.
– means that the variable is not included in the estimation.

ing through the country in question on the way to another (third) country.

As is shown in Table 3, the inclusion of the “world trade” variable appears to have improved the ADL models in terms of $\overline{R}^2$s. Individually, the "world trade" variable appears to be significant in all routes. It is also the more important predictor, taking into account the estimated coefficients and the fact that the average world trade growth (at 5.75%) is larger than that of any of the individual country impacts. The same variables that have been included in the ADL models (with and without the "world trade" variable) are then used to build up models for the pooled ADL and TVP models (the results of which are not included in the paper, but are available upon request).

4.1. Forecasting performance

In this sub-section, we compare the forecasting performances of various models, including the ADL, Pooled ADL, TVP and PcGive automatic models, with all of these models also supplemented by including the "world trade" variable. The unconditional VAR model is also included in the comparison.[3] In addition, the two naïve models, along with the exponential smoothing and AR(3) models, provide benchmarks. Four error measures have been used: Root Mean Square (Percentage) Error (RMSE), Geometric Root Mean Square Error (GRMSE), Mean Absolute Scaled Error (MASE), and Geometric Mean Relative Absolute Error (GRelAE).

Table 4 presents the evaluation of the one-step-ahead forecasting performances of various models for two of the error statistics, RMSE and one of the relative measures, GRelAE.[4] Table 6 summarises the performance across all lead times and error measures. Before coming to the details, the broad picture is as follows:

- The addition of the "world trade" variable consistently improves forecasting performance for the econometric models.
- The ADL models consistently perform better than the univariate alternatives of exponential smoothing.
- The subjectively specified ADL models (using the heuristics described above) perform better than the PcGive automatic algorithmic approach.
- Pooling adds little when the "world trade" variable is included in the model, and is less effective than including the "world trade" variable alone.
- The VAR models are out-performed by the ADL model (with world trade), with the differences often being quite substantial (based on most of the error measures). This is not surprising, given the additional information available to the ADL model. More surprising is the fact that the VAR model performs comparatively better for leads of 2 and 3, despite the fact that the VAR is using the poorly predicted lead 1 lag.[5]
- The error measures all tell broadly the same story (with rank correlations around 0.8); however, a complete picture requires a full analysis of where the differences between measures arise. The differences are most pronounced when comparing RMSE and GRelAE.

In more detail, the GRelAE measures show NM2 demonstrating a substantial improvement over NM1, the zero growth forecast (illustrated in Table 4 by lines 2a and 2b compared to 1a and 1b, respectively). The RMSE measures show that there is little difference between the univariate autoregressive forecasts and the exponential smoothing results (lines 3 and 4). However, the RMSE and the GRelAE here tell different stories. The adoption of a more robust (to outliers) error measure, standardised by the forecasting difficulty of the series, suggests that NM2 performs more strongly (a point we pick up on later in discussing the overall results).

Next to be compared is the performance of the ADL models. The median RMSE shows that the ADL model (line 5a) generally performs better than the NM1, NM2 and AR(3) models, but the evidence is conflicting for GRelAE. In fact, the median of GRelAE is larger than that of NM2, AR(3) and exponential smoothing, though it is smaller than NM1. The explanation may lie in the sometimes unconvincing model specification for most of the estimated ADL models (see Table 2, particularly with regard to the UK–Sweden and UK–USA routes).

[3] Based on the AIC criterion, only the USA–Canada route includes lags of up to 3; for all other routes, only one period lagged variables are included as regressors.

[4] Two of the four error measures are presented in Table 4 here. This is to save space. The conclusions based on the four measures are all broadly in line with the discussion in the text. The full results are available upon request.

[5] We thank a referee for pointing this out.

Table 4
Comparative performance of one-year-ahead forecasts of air traffic models, measured by RMSE and GRelAE (Evaluation database: 1992–2002).

Models	Routes										Median	
	UK–Germany		UK–Sweden		UK–Italy		UK–USA		UK–Canada			
	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank	RMSE	Rank
1a. NM1	6.85	(8)	10.35	(10)	9.21	(11)	8.53	(11)	7.06	(10)	8.53	(10)
2a. NM2	7.55	(10)	11.27	(11)	9.85	(12)	9.97	(12)	6.62	(9)	9.85	(12)
3a. Exponential smoothing	6.78	(7)	7.27	(5)	6.52	(6)	7.19	(6)	5.77	(5)	6.78	(7)
4a. AR(3)	6.41	(6)	7.16	(4)	6.82	(7)	7.95	(10)	5.54	(3)	6.82	(8)
5a. ADL	5.01	(5)	7.91	(7)	4.51	(2)	6.38	(4)	5.85	(6)	5.85	(4)
6a. Pooled ADL	4.18	(3)	7.59	(6)	4.23	(1)	6.52	(5)	5.70	(4)	5.70	(3)
7a. PcGive auto	8.87	(12)	10.22	(9)	7.42	(8)	7.48	(7)	9.56	(12)	8.87	(11)
8a.TVP	4.36	(4)	8.05	(8)	5.97	(5)	7.68	(9)	6.21	(7)	6.21	(5)
9a. ADL +“World trade” variable	3.53	(1)	4.92	(3)	4.61	(3)	5.06	(2)	5.48	(2)	4.92	(2)
10a. Pooled ADL+“World trade” variable	4.10	(2)	4.90	(1)	4.80	(4)	4.61	(1)	5.35	(1)	4.80	(1)
11a. PcGive Auto +“World trade” variable	6.91	(9)	19.69	(12)	7.58	(9)	7.58	(8)	8.02	(11)	7.58	(9)
12a. TVP +“World trade” variable	8.40	(11)	4.92	(2)	7.71	(10)	6.04	(3)	6.52	(8)	6.52	(6)
13a. VAR	8.61	(12)	7.96	(8)	5.42	(6)	6.50	(5)	7.55	(11)	7.55	(9)
14a. VAR +“World trade” variable	9.18	(14)	8.01	(9)	5.10	(5)	6.86	(7)	7.66	(12)	7.66	(11)
	GRelAE	Rank	GRelAE	Rank	GRelAE	Rank	GRelAE	Rank	GRelAE	Rank	MdRelAE	Rank
1b. NM1	100	(14)	100	(10)	100	(14)	100	(13)	100	(14)	100	(14)
2b. NM2	54.5	(9)	97.5	(8)	50.4	(8)	39.4	(2)	44.4	(1)	50.4	(2)
3b. Exponential smoothing	56.2	(10)	52.8	(2)	37.9	(1)	79.3	(10)	58.3	(3)	56.2	(7)
4b. AR(3)	37.6	(3)	52.7	(1)	44	(3)	75.1	(8)	76.7	(11)	52.7	(5)
5b. ADL	40.3	(5)	108	(13)	47.7	(6)	97.5	(12)	69.1	(10)	69.1	(12)
6b. Pooled ADL	41.3	(6)	101	(12)	39.7	(2)	85.2	(11)	58.4	(4)	58.4	(8)
7b. PcGive auto	76.5	(13)	87.8	(7)	56.7	(11)	139	(14)	85.9	(12)	85.9	(13)
8b.TVP	62.7	(11)	109	(14)	47.7	(5)	76.6	(9)	68	(7)	68	(11)
9b. ADL +“World trade” variable	45.6	(8)	66.6	(5)	49.1	(7)	42.8	(4)	58	(2)	49.1	(1)
10b. Pooled ADL +“World trade” variable	21	(1)	65.9	(3)	47.2	(4)	50.9	(6)	60.7	(6)	50.9	(3)
11b. PcGive Auto +“World trade” variable	65.2	(12)	87.6	(6)	59.2	(12)	56.5	(7)	86	(13)	65.2	(10)
12b. TVP +“World trade” variable	39	(4)	66.4	(4)	61.5	(13)	40.6	(3)	60.4	(5)	60.4	(9)
13b. VAR	44.8	(7)	98.5	(9)	55.7	(10)	50.2	(5)	68.2	(8)	55.7	(6)
14b. VAR +“World Trade” variable	35.4	(2)	100	(11)	51.2	(9)	33.3	(1)	68.7	(9)	51.2	(4)

Notes: The values in parentheses indicate the ascending ranks of the RMSE and GRelAE, respectively.

In addition, both error measures show that the ADL model (lines 5a and 5b) performs better than the automatically specified PcGive automatic model (lines 7a and 7b). However, pooling data from different countries leads to improved accuracy for all error measures (lines 6a and 6b). Despite our initial hypothesis, the TVP models (lines 8a and 8b) perform poorly.

The results of including the “world trade” variable in the ADL models (lines 9a and 9b) show that in 7 out of 10 models, there is a decrease in the two error measures shown. The ADL model outperforms the PcGive automatic model in terms of RMSE when the “world trade” variable is included (lines 10a and 11a), but the PcGive automatic model appears to outperform the ADL model based on GRelAE (lines 10b and 11b).

Generally the pooled models with world trade data show an improvement over unpooled models, regardless of which error measure is used. This is similar

Table 5
Forecast performance of ADL vs TVP vs benchmark alternatives: Harvey et al.'s (1997) test, p-values.

RMSE	Lead 1	Lead 2	Lead 3
ADL with "World trade" variable vs. NM2	0.000	0.000	0.000[a,b]
ADL with "World trade" variable vs. Exponential smoothing	0.001	0.002	0.000
ADL with "World trade" variable vs. AR(3)	0.001	0.013	0.000
ADL with "World trade" variable vs. Pooled ADL with "World trade" variable	0.017	0.000[c]	0.001[c]
ADL with "World trade" variable vs. PcGive automatic	0.000	0.001	0.000
ADL with "World trade" variable vs. TVP with "World trade" variable	0.001	0.005	0.040
TVP with "World trade" variable vs. AR(3)	0.010	0.003	0.010[a]
TVP with "World trade" variable vs. PcGive automatic	0.040	0.010	0.043

[a] Omitting the UK–Canada route, where the asymptotic variance could not be calculated.
[b] Omitting the UK–Germany route, where the asymptotic variance could not be calculated.
[c] Omitting the UK–Sweden route, where the asymptotic variance could not be calculated.

to the results of Garcia-Ferrer et al. (1987). The TVP model shows no improvement over the fixed parameter models (lines 12a and 12b).

Adding in the world trade variable is critical to the success of the ADL model. Once included, pooling adds little to the forecast accuracy, which is an intuitively reasonable observation, as the two approaches have a common intention of taking into account common features in air travel across counties.

Finally, a key question is the performance of the econometric models compared to the unconditional VAR model and the simple univariate extrapolative alternatives. If the RMSE is used as the error measure, the ADL model with world trade (from the study by Garcia-Ferrer et al., 1987, of the expected best performer) outperforms exponential smoothing (the best performer in the various univariate forecasting competitions) for all five countries. The VAR model performs no better than the simpler AR(3) alternative (lines 4a, 13a and 14a). However, if the relative absolute error measure (GRelAE) is used, the result is too close to call (versus either the VAR or the AR(3), lines 4b and 14b). This leads to two conclusions:

(1) Before a method can be identified as 'best', the error measure that fits with the decision problem needs to be specified.

(2) With this particular data set, despite the results being averaged over effectively 55 data points, the variability in performance effectively dominates. Note however that, when measured by the RMSE, the difference could be viewed as substantial, at 4.80 compared to 6.78 (approximately 2 percentage points).

Following on from this discussion of the importance of the differences in performance, the root mean squared errors (and geometric mean squared errors) can be tested for significant differences by following Harvey, Leybourne, and Newbold's (1997) approach. By pooling the approximate p-values from the test statistics for each country, the p-values of the RMSE and, using the log (squared error), the GRMSE can be estimated. The RMSE results are shown in Table 5. A number of comparisons were carried out, comparing the ADL model including the "world trade" variable with the Naïve model 2, the Exponential smoothing model, the AR(3) model, the TVP model including the "world trade" variable, and the PcGive automatic-based ADL model including the "world trade" variable. The results show that for lead 1, the null hypothesis of the equality of RMSEs between the ADL and other models mentioned above can be rejected, with the same result for the null hypothesis of the equality of GRMSEs, despite the variability in the individual country results. We also applied these tests for both the RMSE and GRMSE between the TVP and the PcGive automatic-based ADL models, which again resulted in the rejection of the null. For leads 2 and 3, the same conclusions can be reached. The overall performance of the ADL model proves to be significantly better than that of the benchmark alternatives, but, following the arguments of Armstrong (2007), more importantly, the differences are relatively large for some of the error measures, and are in accord with prior research. However, they are perhaps not of economic significance, in that the decisions as to such issues as route capac-

Table 6
Comparative performances of one-, two- and three-year-ahead forecasts of air traffic models, based on the RMSE, GRMSE, MASE and GRelAE error measures.

Model	RMSE							MASE							Rank over all leads + error measures
	Lead 1		Lead 2		Lead 3		Overall rank	Lead 1		Lead 2		Lead 3		Overall rank	
1 NM1	8.53	(12)	7.26	(12)	7.47	(8)	(11)	93.2	(14)	92.7	(14)	98.1	(14)	(14)	(14)
2 NM2	9.85	(14)	8.88	(14)	9.10	(14)	(14)	82.5	(12)	76.1	(12)	88.4	(13)	(13)	(12)
3 Exponential smoothing	6.78	(7)	7.12	(10)	7.79	(13)	(10)	65.8	(5)	57.8	(4)	63.4	(6)	(5)	(6)
4 AR(3)	6.82	(8)	6.78	(6)	7.60	(11)	(8)	52.8	(4)	61.4	(6)	51.1	(3)	(3)	(3)
5 ADL	5.85	(4)	6.93	(8)	6.86	(7)	(7)	71.8	(8)	65.8	(9)	69.4	(9)	(9)	(11)
6 Pooled ADL	5.70	(3)	6.80	(7)	6.85	(6)	(5)	68.9	(6)	65.6	(7)	70.8	(12)	(8)	(4)
7 PcGive auto	8.87	(13)	8.26	(13)	7.70	(12)	(13)	76.2	(10)	77.1	(13)	69.2	(8)	(12)	(13)
8 TVP	6.21	(5)	6.02	(4)	6.70	(5)	(4)	72.6	(9)	66.8	(10)	69.9	(11)	(11)	(10)
9 ADL +"World trade" variable	4.92	(2)	4.96	(2)	5.20	(2)	(2)	51.3	(3)	45.1	(2)	46.4	(1)	(2)	(1)
10 Pooled ADL +"World trade" variable	4.80	(1)	4.80	(1)	4.98	(1)	(1)	43.6	(1)	44.9	(1)	47.1	(2)	(1)	(2)
11 PcGive Auto +"World trade" variable	7.58	(10)	6.08	(5)	6.39	(3)	(6)	88.9	(13)	67.6	(11)	62.3	(5)	(10)	(9)
12 TVP +"World trade" variable	6.52	(6)	5.67	(3)	6.52	(4)	(3)	50.1	(2)	65.7	(8)	69.7	(10)	(6)	(7)
13 VAR	7.55	(9)	7.10	(9)	7.55	(9)	(9)	76.5	(11)	58.0	(5)	63.6	(7)	(7)	(8)
14 VAR +"World trade" variable	7.66	(11)	7.12	(11)	7.58	(10)	(11)	70.7	(7)	57.5	(3)	56.0	(4)	(4)	(4)
	GRMSE							GRelAE							
	Lead 1		Lead 2		Lead 3		Overall rank	Lead 1		Lead 2		Lead 3		Overall rank	
1 NM1	5.78	(14)	5.20	(13)	5.49	(14)	(14)	100	(14)	100	(14)	100	(14)	(14)	
2 NM2	3.62	(10)	5.35	(14)	4.59	(13)	(13)	50.4	(2)	80.4	(12)	87.9	(13)	(9)	
3 Exponential smoothing	3.08	(3)	2.74	(4)	2.86	(5)	(3)	56.3	(7)	55.7	(4)	51.7	(3)	(4)	
4 AR(3)	3.13	(4)	2.90	(6)	2.73	(3)	(4)	52.8	(5)	61.4	(7)	51.1	(2)	(4)	
5 ADL	3.42	(7)	3.94	(11)	4.20	(12)	(11)	69.1	(12)	74.5	(11)	81.0	(12)	(12)	
6 Pooled ADL	2.95	(1)	2.80	(5)	2.76	(4)	(1)	58.5	(8)	59.4	(6)	59.4	(6)	(6)	
7 PcGive auto	4.54	(13)	5.05	(12)	4.10	(11)	(12)	85.9	(13)	90.6	(13)	69.2	(10)	(13)	
8 TVP	3.44	(8)	3.59	(9)	3.62	(7)	(8)	68.0	(11)	65.4	(8)	69.8	(11)	(11)	
9 ADL +"World trade" variable	3.06	(2)	2.60	(3)	3.24	(6)	(2)	49.1	(1)	44.6	(2)	56.2	(5)	(2)	
10 Pooled ADL +"World trade" variable	3.21	(6)	3.11	(7)	3.94	(8)	(7)	50.9	(3)	55.6	(3)	55.6	(4)	(3)	
11 PcGive Auto +"World trade" variable	4.25	(12)	3.41	(8)	3.98	(9)	(10)	65.2	(10)	57.0	(5)	62.3	(7)	(7)	
12 TVP +"World trade" variable	3.19	(5)	3.86	(10)	4.09	(10)	(9)	60.4	(9)	68.8	(10)	65.4	(8)	(9)	
13 VAR	3.61	(9)	2.48	(2)	2.01	(2)	(4)	55.7	(6)	68.2	(9)	68.2	(9)	(8)	
14 VAR +"World trade" variable	3.63	(11)	2.33	(1)	1.82	(1)	(4)	51.2	(4)	40.3	(1)	41.2	(1)	(1)	

Notes: The values in parentheses indicate the ascending rankings for the different error measures. Overall rankings by error measures and summarised across error measures are also shown.

ity and landing slots are unlikely to be sensitive to the relatively small absolute improvements.

Table 6 provides summary information on one-, two- and three-year-ahead forecasts, and gives an overall summary using all four error measures. Overall, the results for two- and three-year-ahead forecasts are similar to the results for the one-year-ahead forecasts shown in Table 4. The two univariate methods (exponential smoothing and the AR(3)) are overall third and fourth compared to the subjectively specified ADL models that included world trade. The differences are small (around 10%), apart from those measured through the RMSE.

To check the models' sensitivity, we re-evaluated models with the last two data points omitted, in order to exclude the effect of September 11, 2001, and statistical tests show that the results do not change significantly. The model ranking conclusions also remain the same. Tables similar to Tables 4 and 6 are available upon request. These results permit us to answer the question of the effects of September 11. Using the best model (that is, ADL with the "world trade" variable included) estimated with data up to 2000 gives errors for 2001 and 2002 with a median value which is significantly negative at the 5% level. By 2002, the indirect effects of the catastrophe on trade and income were coming through, so the overestimate was less (based on the realised values of the explanatory variables). If the average growth in world trade had continued in 2002, we estimate that air travel would have been 8.9% higher. These results are consistent with the forecasts to the end of 2002 presented by Blunk et al. (2006).

5. Discussion and conclusions

In this research, several econometric models have been specified and estimated for air passenger traffic demand between the UK and five selected countries. They included three model types that have proved effective in earlier studies, and one goal of this research was to see whether these results could be reproduced in a different context. The model types were autoregressive distributed lag models that included a common 'world' variable, a system model that provided pooled estimates, and time-varying parameter models.

The demand models developed were then evaluated for their (conditional) out-of-sample forecasting accuracy. The results proved robust over the three-year lead time. The one-, two- and three-year-ahead ex-post forecasts were generated and compared with the benchmark models, namely naïve model 1 (NM1) and naïve model 2 (NM2), as well as two established univariate alternatives, an AR(3) model and an exponential smoothing model.

Various error criteria were used in evaluating model performances across longer lead times than is typical. Different error measures have tended to produce different results in forecast evaluations, and therefore any discussion of comparative forecasting performances must be supported by evidence obtained from several error measures. Interestingly, an analysis of the error in the cumulative growth forecast for 3 years ahead showed an increased dominance of the 'ADL model with world trade'. This adds support to the principle that econometric models perform comparatively better when there are large changes in the explanatory variables.

Regarding the main hypothesis that econometric models will outperform univariate benchmarks, the ADL models with the "world trade" variable consistently outperformed both the naïve models and the two time series methods. However, only for the UK–USA and UK–Canada routes did the 'ADL model with world trade' outperform the AR(3) model when all of the error measures and lead times were considered, despite their aggregate strong performance, as shown in Table 6. The UK–Canada model has the highest explanatory power with no autoregressive component. The remaining comparisons offer further examples of a pure time series model appearing to be more accurate than a causal model. A tentative explanation for this is that the proposed causal model could be better specified with the inclusion of other important drivers, and in particular measures of structural change in the market, such as increased price competition. An alternative rationalisation lies in the possible effects of model complexity on estimation reliability in relatively small samples, which can lead to autoregressive models outperforming even well-specified structural models, as was the case here (Favero & Marcellino, 2005).

A subsidiary hypothesis concerned the relative performance of the subjectively specified ADL model compared to that of the automatic specification delivered by PcGive. The ADL models (with or without the "world trade" variable) proved better performers than PcGive automatic, contradicting our first hypothesis as to the benefits of removing subjectivity in model building. The difference arises from the relative parsimony (in the larger samples) of PcGive automatic models in attempting to avoid including spurious relationships. With the smaller sample sizes there was evidence of multicollinearity in the models, as specified by the Pc-Give automatic based models. This result emphasises the need for further research examining the differences between automatically specified models and their subjectively specified alternatives.

Simultaneously pooling the data and estimating the models improved the forecasting performance, although the differences were slight. This confirms the conclusions of Garcia-Ferrer et al. (1987) and Zellner et al. (1991), who found that forecasting models recognizing contemporaneous co-variation and using the seemingly unrelated regression (SUR) approach showed improvements over models estimated individually.

The time-varying parameters (TVP) model failed to show the expected improvements over fixed parameters models, in conflict with the principle laid down by Allen and Fildes (2001), that time-varying parameter models are most valuable when the appropriate model structure is not well-understood, or unobserved variables are affecting its structure. This result is also inconsistent with findings in tourism demand forecasting studies (Li & Song et al., 2006). An examination of the parameter variation in the TVP models in the forecast period shows little variation for most countries, but the patterns look much the same as those of Li and Wong et al. (2006). Only Italy showed a predictable parameter drift, a condition that would suggest that TVP models might outperform their fixed parameter equivalents, although there is no evidence of this in the results. This apparent contradiction suggests a need for further detailed research.

The results depend on the chosen error measure, and although the rankings are positively correlated, there remains a clear difference between the relative measures (MASE and GRelAE) and the RMSE and GRMSE, which are not standardised. An investigation into the error distributions (absolute and relative) revealed a small number of outliers which have affected the country-level results, changing the overall rankings. While arguments such as the robustness to outliers (Fildes, 1992) and data characteristics (Hyndman & Koehler, 2006) are important and argue for the use of relative measures, the choice in practice rests with the users' preferences. The findings from this research show that when reporting empirical results, a range of measures are needed which incorporate natural metrics that fit the (often implicit) decision problem and associated loss function, as well as one of these relative metrics.

September 11, 2001, naturally had a major negative impact on the airline industry. However, 2002 data showed that the recovery process was already underway, although it was not back to its long term growth path. Empirical findings for the US from Lai and Lu (2005), as well as those for the UK, Germany and Australia from Njegovan (2006), also indicate that shocks to air passenger traffic are largely transitory, and do not, in general, merit the revision of forecasts over a long horizon. To check the models' sensitivity, we re-estimated ADL models with the world trade variable by including a dummy variable which equals 1 for the last two data points, in order to exclude the effect of September 11. The results show that the dummy variable is insignificant for all 5 models, and the coefficients on other variables appear to be stable. If we include the dummy variable from the beginning when simplifying the ADL model, only the model for the UK–Canada route is changed, with more variables now included in the model and the dummy variable appearing to be significant at the 10% level.

In part, this study has attempted to replicate and extend aspects of the Garcia-Ferrer et al. (1987) study of GDP growth. The research has confirmed the difficulties of such replications and the choices made by earlier researchers. We have therefore attempted to be explicit in Section 3 about our model building and the software we have used. For example, the choice of priors in the TVP modelling proved important in Garcia-Ferrer et al.'s study. Our use of standard priors, as discussed in Section 3, led to a poorer performance than our preliminary models had suggested. One area of potential importance in such replications is the choice of data and error measures. Here we have examined the effects of omitting 2001/2 because of the September 11 effect, and have also considered a variety of error measures. However, different countries and different measures of price (in particular) might well lead to different conclusions. The importance of these replications, as Hubbard and Vetter (1996) remarked, is that they cumulatively lead to a greater understanding of the effectiveness of alternative [forecasting] models. One-off studies, however expertly carried out, can do no such thing.

In summary, the results of this research show that employing appropriately specified structural econometric methods leads to an improved forecasting performance for air traffic, compared to non-structural alternatives, although our comparison is conditional, and therefore favours the structural methods. This adds further evidence to the weak conclusions drawn by Allen and Fildes (2001) as to the benefits of econometric models compared to their univariate alternatives. The absolute differences between the two are small, however, and arguments for simplicity suggest that a simple univariate model would be adequate for most users. The use of a VAR model here adds little. Our key subsidiary hypotheses, laid out in the introduction, have also found support in this replication. However, the adoption of a TVP approach did not help, despite the apparent structural changes in the market. Nor did the adoption of an automatic model-building approach for specifying the ADL model. These results, together with the structural changes seen in the UK travel market since 2002, and in particular the rise of the low cost carriers, may well require a different approach to modelling, based on forecasting the two competing market segments separately.

Acknowledgements

Mohd Alias Lazim contributed to early drafts of this research. The authors would like to thank an anonymous referee, whose persistence has enhanced and re-directed the research focus of this paper.

Appendix

Income

Income is personal disposable income. Data are obtained from the following sources: Yearbook of National Account Statistics for data on personal disposable income for the UK, The Statistical Abstract of the United States for data on personal disposable income for the US, and OECD publications for data on personal disposable income for the rest of the countries in the study.

Price

There are several difficulties that need to be addressed when incorporating price data into the model. The main difficulty is that there is no single price for flights between any two country-points, because there are several possible points connecting the UK and country i. Therefore, typical points have to be selected in the UK and in country i, such that the price of travel between these typical points can act as a representative sample of prices of travel between the UK and country i. In this research, a point is chosen on the basis of its importance as a gateway into/out of a country. The points selected for each country are as follows:

Country	Point
UK	London
Germany	Frankfurt
Sweden	Stockholm
Italy	Rome
USA	New York
Canada	Vancouver

For example, the price for the UK–Germany route is the return fare between London and Frankfurt, to reflect the demand for the UK's and Germany's origin passengers. Similar procedures are used on all other routes.

In this research the following points are considered when choosing the fares to represent prices. The best rationale for choosing the economy fare as an indicator of price levels between the UK and country i is that it is the largest contributor of passenger sales in any flight. For example, of all British Airway's passengers, three-quarters choose to fly economy, and of all Virgin Atlantic's passengers, only 10%–15% fly Upper Class (The Sunday Times, 28 March 1993). For model construction purposes, a single figure is to be used to represent the price for the year. The average of prices in May and November across all published economy fares sold by major airlines is chosen to represent the price. These two months are selected to avoid the effect of seasonal change in fares as much as possible. Finally, the representativeness of the sample values of the price is based on an assumption that there are no marked differences between the fares of different carriers operating on similar routes, which relate to IATA's objectives. (Part of the function of IATA is to act as a regulatory body on price agreements between countries.) The price data are obtained from The ABC

World Airways Guide (various years) for 1961–1998 and CAA (UK) for 1999–2002.

Trade

The inclusion of an individual trade variable is expected to reflect the strength of the economic relationship between the UK and country i. This is based on the assumption that the greater the opportunities available to business communities in the two countries, the more people we would expect to travel between the two countries for business purposes. The trade data are obtained from the following sources: The UN Yearbook of International Trade Statistics, Vol. 1, for data on Trade by Country, and The International Financial Statistics Yearbook (UN) for import/export data and data on market/par rates.

Data transformation

The explanatory variables included in the model are measured in different units. It is therefore important to transform them to make data comparable. The transformation procedures are as follows:

- The transformation of trade figures begins by deflating, using the UK's unit value of exports and unit value of imports.

$$\text{Trade}_i = ((\text{US\$ Exports}_i / XU) * 100) + ((\text{US\$ Imports}_i / MU) * 100),$$

  where XU = UK's unit value of exports; MU = UK's unit value of imports; US\$ Exports_i = exports to country i in US\$, and US\$ Imports_i = imports from country i in US\$.
- To obtain the real price on each route and the real personal disposable income, the respective country's Consumer Price Index (CPI) is used as the deflator. Since both fares and personal income are given in terms of the local currency, the deflated figures are then converted to US\$ using the market exchange rates provided in the International Financial Statistics Yearbook.

Weighting procedure

Weights are used to reflect the influences of the factors at both ends of the route. In this research, the weights used are the proportion of arrivals and departures into and out of the UK of residents of country i and the UK, respectively (see the table below). These proportions are calculated based on the average of the number of passengers over a ten-year period, broken down by nationality, arriving in/departing from the UK. For example, visitor arrivals from Germany are assumed to be German nationals, and visitors departing from the UK to go to Germany are assumed to be UK residents.

A similar procedure is applied by Rodriquez (1981), who assigns a constant weight of 1/3 for Mexico and 2/3 for the US, to reflect the proportion of travellers of each nationality in the forecast model of international traffic between Mexico and the US. The Association of European Airlines (AEA) uses the proportion of sales in each country to derive the weights. In this study, the weights (α_1 and α_2) are calculated based on the average over a ten-year period (1982–1991) of the number of visits abroad made by UK residents and the number of overseas visitors to the UK by air, as published in Business Monitors (various years) (CSO) on Overseas Travel and Tourism. The following are the weights obtained by route:

ROUTE	α_1 [a]	α_2 [b]
UK–Germany	0.50	0.50
UK–Sweden	0.32	0.68
UK–Italy	0.65	0.35
UK–USA	0.38	0.62
UK–Canada	0.36	0.64

[a] The proportion of travellers of UK origin flying to respective country i.

[b] The proportion of travellers of respective origin country i travelling to the UK by air.

Then the weighted personal disposable income and price were calculated as follows:

$$\text{WI} = \alpha_1 \text{IUK}_t + \alpha_2 \text{IC}i_t,$$

where WI is the weighted average of personal disposable income of the UK and country i in year t, IUK_t is the personal disposable income for the UK in year t, and $\text{IC}i_t$ is the personal disposable income for country i in year t.

$$\text{WP} = \alpha_1 \text{LC}i_t + \alpha_2 \text{C}i\text{L}_t,$$

where WP is the weighted average of the return fare for the UK–country i route in year t, $\text{LC}i_t$ is the median of the return fare for the UK–country i route in year t, and $\mathrm{C}i\mathrm{L}_t$ is the median of the return fare for the country i–UK route in year t.

References

ABC World Airways Guide, (various years). Dunstable, UK, various issues.

Abed, S. Y., Ba-Fail, A. O., & Jasimuddin, S. M. (2001). An econometric analysis of international air travel demand in Saudi Arabia. *Journal of Air Transport Management*, *7*, 143–148.

Allen, P. G., & Fildes, R. (2001). Econometric forecasting: Strategies and techniques. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers.

Allen, P. G., & Fildes, R. (2005). Levels, differences and ECMs — Principles for improved econometric forecasting. *Oxford Bulletin of Economics and Statistics*, *67*, 881–904.

Anderson, J. E., & Kraus, M. (1980). Quality of service and the demand for air travel. *The Review of Economics and Statistics*, *63*, 533–540.

Armstrong, J. S. (2007). Significance tests harm progress in forecasting. *International Journal of Forecasting*, *23*, 321–327.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. (2008). *The tourism forecasting competition*. Dept. Business and Econometrics, Monash University, Working paper 10/08. http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2008/wp10-08.pdf.

Blunk, S. S., Clark, D. E., & McGibany, J. M. (2006). Evaluating the long-run impacts of the 9/11 terrorist attacks on US domestic airline travel. *Applied Economics*, *38*, 363–370.

Business Monitors (various years). Central Statistical Office.

Charemza, W. W., & Deadman, D. F. (1997). *New directions in econometric practice*. Edward Elgar.

Clements, M. P., & Hendry, D. F. (1998). *Forecasting economic time series*. Cambridge: Cambridge University Press.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*, 1668–1674.

Du Preez, J., & Witt, S. F. (2003). Univariate versus multivariate time series forecasting: An application to international tourism demand. *International Journal of Forecasting*, *19*, 435–451.

Favero, C. A., & Marcellino, M. (2005). Modelling and forecasting fiscal variables for the Euro area. *Oxford Bulletin of Economics and Statistics*, *67*, 755–783.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, *8*, 81–98.

Fildes, R. A., & Ord, J. K. (2002). Forecasting competitions — Their role in improving forecasting practice and research. In M. Clements, & D. Hendry (Eds.), *A companion to economic forecasting* (pp. 322–353). Oxford: Blackwell, ISBN: 0-6312-1569-7.

Fridström, L., & Thune-Larsen, H. (1989). An econometric air travel demand model for the entire conventional domestic network: The case of Norway. *Transportation Research Part B*, *23*, 213–223.

Garcia-Ferrer, A., Highfield, R. A., Palm, F., & Zellner, A. (1987). Macroeconomic forecasting using pooled international data. *Journal of Business and Economic Statistics*, *5*(1), 53–67.

Goodrich, R. L. (2000). The ForecastPro methodology. *International Journal of Forecasting*, *16*, 533–535.

Grubb, H., & Mason, A. (2001). Long lead-time forecasting of UK air passengers by Holt–Winters' methods with damped trend. *International Journal of Forecasting*, *17*, 71–82.

Harvey, A. C. (1989). *Forecasting structural time series models and the Kalman Filter*. Cambridge: Cambridge University Press.

Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, *13*, 281–291.

Hendry, D. F. (1986). Econometric modelling with co-integrated variables: An overview. *Oxford Bulletin of Economics and Statistics*, *48*, 201–211.

Hendry, D. F., & Krolzig, H.-M. (2001). *Automatic econometric model selection using PcGets 1.0*. Timberlake Consultants Ltd.

Hubbard, R., & Vetter, D. E. (1996). An empirical comparison of published replication research in accounting, economics, finance, management, and marketing. *Journal of Business Research*, *35*, 153–164.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*, 679–688.

Ippolito, R. A. (1981). Estimating airline demand with quality of service variables. *Journal of Transport Economics and Policy*, *15*(1), 7–15.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H., & Lee, T.-C. (1985). *The theory and practice of econometrics*. New York: Wiley.

Jorge-Calderon, J. D. (1997). A demand model for scheduled airline services on international European routes. *Journal of Air Transport Management*, *3*, 23–35.

Kaemmerle, K. C. (1991). Estimating the demand for small community air service. *Transportation Research Part A*, *25*, 101–112.

Koopman, S. J., Shephard, N., & Doornik, J. A. (1999). Statistical algorithms for models in state space using SsfPack 2.2. *Econometric Journal*, *2*(1), 107–160.

Lai, S. L., & Lu, W. L. (2005). Impact analysis of September 11 on air travel demand in the USA. *Journal of Air Transport Management*, *11*(6), 455–458.

Li, G., Song, H., & Witt, S. F. (2006). Time varying parameter and fixed parameter linear aids: An application to tourism demand forecasting. *International Journal of Forecasting*, *22*, 57–71.

Li, G., Wong, K. F., Song, H., & Witt, S. F. (2006). Tourism demand forecasting: A time varying parameter error correction model. *Journal of Travel Research*, *45*, 175–185.

Lütkepohl, H. (1993). *Introduction to multiple time series analysis* (2nd ed.). Springer.

Matsumoto, H. (2004). International urban systems and air passenger and cargo flows: Some calculations. *Journal of Air Transport Management*, *10*(4), 239–247.

National Statistics Office (2002). *Travel Trends 2002. A report on the national passenger survey*. Her Majesty's Stationary Office, Norwich.

Njegovan, N. (2006). Are shocks to air passenger traffic permanent or transitory? Implications for long-term air passenger forecasts for the UK. *Journal of Transport Economics and Policy*, *40*(2), 315–328.

Pagan, A. R. (1987). Three econometric methodologies: A critical appraisal. *Journal of Economic Surveys*, *1*, 3–24.

Pagan, A. R. (1999). The Tilburg experiments: Impressions of a drop-out. In J. R. Magnus, & M. S. Morgan (Eds.), *Methodology and tacit knowledge* (pp. 369–374). Chichester, UK: Wiley.

Quandt, R. E., & Baumol, W. J. (1966). The demand for abstract transport models: Theory and measurement. *Journal of Regional Science*, *6*(2), 13–26.

Riddington, G. L. (1993). Time varying coefficient models and their forecasting performance. *Omega*, *21*, 573–583.

Rodriquez, L. L. (1981). Traffic forecast for planning purposes. Presented at an *ICAO regional workshop on aviation forecasting and economic planning*, Mexico City.

Swan, W. (2002). Three mistakes we made in forecasting. Presented at the *international symposium on forecasting*, Dublin.

Young, K. H. (1972). A synthesis of time series and cross-section analyses: Demand for air transportation service. *Journal of the American Statistical Association*, *67*, 560–566.

Zellner, A., Hong, C., & Min, C. (1991). Forecasting turning points in international output growth rates using Bayesian exponentially weighted auto-regression, time-varying parameter and pooling techniques. *Journal of Econometrics*, *49*, 275–304.

Zellner, A., & Min, C. (1993). Bayesian and non-Bayesian methods for combining models and forecasts, with applications to forecasting international growth rates. *Journal of Econometrics*, *56*, 89–118.

Zellner, A., & Palm, F. C. (2004). *Structural econometric modelling, time series analysis (SEMSTA) approach*. Cambridge: Cambridge University Press.

Robert Fildes is Distinguished Professor of Management Science in the School of Management, Lancaster University, and Director of the Lancaster Centre for Forecasting. He has a mathematics degree from Oxford and a Ph.D. in statistics from the University of California. He was co-founder of the *Journal of Forecasting* in 1981 and of the *International Journal of Forecasting* in 1985. He has published widely in journals including both forecasting journals and the *Journal of the Operational Research Society*, *Management Science*, *Decision Support Systems* and *Oxford Bulletin of Economics and Statistics.* His current research interests are concerned with the comparative evaluation of different forecasting methods, the implementation of improved forecasting procedures in organizations and the design of forecasting systems.

Yingqi Wei (Ph.D., Aston University) is Professor in International Business at the University of York. Her current research interests are in technology adoption and creation and local sourcing of multinational enterprises. She has published in various journals including the *Journal of International Business Studies*, *Journal of Business Research*, *Cambridge Journal of Economics* and *World Economy.*

Suzilah Ismail is assistant professor at the Faculty of Information Technology and Quantitative Sciences, UiTM, Shah Alam. She has a Ph.D. in Management Science from Lancaster University. Her major areas of interest include econometric model specification and the value of automatic algorithms for specifying econometric models.