ELSEVIER

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*

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.

© 2009 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

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.

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 in-