Forecasting (aggregate) demand for US commercial air travel

Richard T. Carson[a], Tolga Cenesizoglu[b,c,*], Roger Parker[d]

[a] *Department of Economics, University of California, San Diego, United States*
[b] *Department of Finance, HEC Montreal, Canada*
[c] *CIRPÉE, 3000, Chemin de la Côte-Sainte-Catherine, Bureau 4.348, Montréal (Québec), H3T 2A7, Canada*
[d] *Virtual Minds, SA, Switzerland*

Abstract

We analyze whether it is better to forecast air travel demand using aggregate data at (say) a national level, or to aggregate the forecasts derived for individual airports using airport-specific data. We compare the US Federal Aviation Administration's (FAA) practice of predicting the total number of passengers using macroeconomic variables with an equivalently specified AIM (aggregating individual markets) approach. The AIM approach outperforms the aggregate forecasting approach in terms of its out-of-sample air travel demand predictions for different forecast horizons. Variants of AIM, where we restrict the coefficient estimates of some explanatory variables to be the same across individual airports, generally dominate both the aggregate and AIM approaches. The superior out-of-sample performances of these so-called quasi-AIM approaches depend on the trade-off between heterogeneity and estimation uncertainty. We argue that the quasi-AIM approaches exploit the heterogeneity across individual airports efficiently, without suffering from as much estimation uncertainty as the AIM approach.
© 2010 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

Keywords: Demand forecasting; Disaggregation; Transportation forecasting; Panel data

1. Introduction

Forecasts of air travel demand are important inputs for a wide variety of economic decisions, including, but not limited to, research and development, airplane design and production planning. For a relatively mature product like air travel, where the interest lies in the aggregate demand, the typical empirical practice is to obtain a national level forecast using aggregate level data when individual market data are not easily accessible (Lehmann & Winer, 2001). However, recent empirical evidence (Bronnenberg, Dhar, & Dube, 2007) has shown that individual markets are much more heterogenous than was thought previously, even for well-known national products which are sold in spatially distinct markets, as is the case for air travel.

An aggregate approach is preferable to a disaggregate approach when the computational/analytical burden of producing forecasts for separate markets is

* Corresponding author at: Department of Finance, HEC Montreal, H3T 2A7, Canada. Tel.: +1 514 340 5668; fax: +1 514 340 5632.

E-mail addresses: cenesizo@yahoo.com, tolga.cenesizoglu@hec.ca (T. Cenesizoglu).

substantial. Furthermore, there are also concerns that in a disaggregate approach, the number of parameters estimated by modeling each market individually quickly becomes large relative to the length of the available time series. On the other hand, the econometric arguments in favor of a disaggregate approach are also fairly strong when disaggregate data are available. Allen and Fildes (2001) review the literature on the advantages of using disaggregate data, one of which is the additional information available due to heterogeneity across individual markets. However, they also argue that the relative performances of aggregate and disaggregate approaches might depend on the specifics of the forecasting exercise.

This paper focuses on exploring the effect of disaggregate information on the accuracy of aggregate air travel demand forecasts. This paper considers the case of US commercial air travel demand, with the objective of predicting the total number of commercial passengers. The data available to us from the US Department of Transportation are the monthly numbers of passengers departing from major US airports between 1990 and 2004. We use individual airport level data for 179 major airports, together with aggregate level data.

We initially consider two extreme approaches. The first follows the FAA's practice of predicting the total number of passengers using a combination of exogenous macroeconomic variables such as population, income, and energy prices in a time series model. The second approach, which we term an AIM (aggregating individual markets) forecast, models the travel demand at the individual market level using exogenous variables specific to that region, then sums the forecasts to produce a forecast of the total number of passengers. Specifically, consider forecasting a variable that is a contemporaneous aggregation of the individual subcomponents at time t:

$$y_t = \sum_{i=1}^{N} y_{it} \quad \text{for } t = 1, 2, \ldots, \tag{1}$$

where y_{it} $(i = 1, 2, \ldots, N)$ are the subcomponents of the aggregate variable y_t. Forecasts of the aggregate variable can be obtained using two different approaches: (1) estimating a reduced-form model for the aggregate variable using aggregate level data, then forming forecasts of the aggregate variable from the estimated model; or (2) estimating a reduced-form model for the subcomponents using individual level data, forming forecasts of the subcomponents from the estimated individual models, then aggregating the subcomponent forecasts to obtain the forecast of the aggregate variable. Not only does the second approach use more information, but also the forecasts at the disaggregate level are readily available, whereas one needs to allocate the aggregate forecast to the individual markets to obtain forecasts at the disaggregate level in the first approach. In this paper, we analyze the out-of-sample forecast performances of these two extreme approaches in forecasting the aggregate variable, along with the performances of other approaches that are between these two extremes.

There has been a revival of interest in the contemporaneous aggregation of disaggregate forecasts to form forecasts of an aggregate variable in both the theoretical and empirical econometrics literature. The empirical literature has focused mainly on forecasting aggregate macroeconomic indices such as Euro-area variables or aggregate US variables by using the information available in the disaggregated subcomponents.[1] In this paper, we not only demonstrate the usefulness of disaggregate information in forecasting an aggregate variable of interest using a new and comprehensive data set, but also discuss different ways of exploiting the heterogeneity across disaggregate level data.

The main contributions of our paper to the existing literature can be summarized as follows. First, applying the two approaches to monthly US air passenger data over the period 1990 to 2002 and then forecasting out-of-sample for the next two years yields a striking contrast: the AIM forecast is far more accurate than the aggregate level forecast. We argue that the performance of the AIM approach depends on the tradeoff between the heterogeneity across markets and the estimation uncertainty due to the number of coefficients estimated. The AIM approach outperforms the aggregate approach, since the forecasting power gained by exploiting heterogenous information across markets dominates the forecasting power lost due to

---

[1] See Benalal, del Hoyo, Landau, Roma, and Skudelny (2004), Espasa, Senra, and Albacete (2002), Fair and Shiller (1990), Marcellino, Stock, and Watson (2003) and Zellner and Tobias (2000).

the estimation of many coefficients. We next consider several variants of AIM (we term the collection of these approaches quasi-AIM), where most of the coefficients are forced to be the same across markets. We find that the quasi-AIM approaches outperform the AIM approach due to the efficient use of heterogeneity across markets with respect to the estimation uncertainty. Quasi-AIMs estimate a much smaller number of parameters relative to AIM, while still using key information about the heterogeneity across markets.

We then consider the forecasting performances of different approaches for three different forecast horizons, short- (1-step-ahead), medium- (6-step-ahead) and long-term (12-step-ahead). This exercise allows us to identify not only the best forecasting approach for different forecast horizons, but also possible reasons for the relative performances in the first exercise. The best performing quasi-AIM approaches in the first exercise are also among the best approaches when we analyze different forecast horizons separately. Independent of the forecast horizon, forecasting approaches which use disaggregate information outperform the aggregate approach. Furthermore, almost all quasi-AIM approaches outperform the AIM approach for every forecast horizon, suggesting that quasi-AIM approaches might be better suited to exploiting the heterogeneity across airports without suffering from as many of the problems associated with estimation uncertainty as the AIM approach.

Our results provide further empirical evidence on the advantages of using disaggregate approaches in forecasting aggregate variables of interest.[2] For example, Giacomini and Granger (2004) show that the

forecasting performance can be improved by imposing *a priori* constraints on the VAR process for the disaggregate variables. They also show that ignoring the impact of spatial correlation, even when it is weak, can lead to highly inaccurate forecasts. Furthermore, Hendry and Hubrich (2006) show that exploiting a common factor structure model in the disaggregate variables might provide better out-of-sample forecasts for the aggregate variable.

The rest of the paper is organized as follows. Section 2 discusses the literature on air travel demand modeling and describes the various different data sources used in our analysis. Section 3 presents the empirical specification and discusses the details of different estimation methods that result in different out-of-sample forecasts of the aggregate air travel demand. Section 4 discusses the forecasting of both the aggregate variable and the explanatory variables. Section 5 presents our forecasting results, and Section 6 concludes.

2. Description of the data

There are several ways of classifying the literature on models of passenger demand for air transportation. In general, the models can be categorized in one of two main subgroups, depending on the choices of the dependent and explanatory variables.

The dependent variable can be either macroscopic or microscopic. In *macroscopic models*, the dependent variable is some aggregate indicator of air travel in a certain country or region. The typical dependent variables are the numbers of flights and passengers, and the passenger revenue miles. *Microscopic models* estimate the air travel demand between two airports, cities or regions. Typical indicators are the passenger traffic in a specific origin-destination (OD) pair route and the number of passengers in each class when there are various tariffs on a route.

The explanatory variables can be divided into two main groups. *Geo-economic factors* are variables such

---

[2] The issue of contemporaneous aggregation dates back to Theil (1954), who argued that the disaggregated approach improves on the model specification of the aggregate variable. Other theoretical papers on the contemporaneous aggregation of disaggregate forecasts include those of Aigner and Goldfeld (1974), Granger (1980, 1987), van Garderen, Lee, and Pesaran (2000), Grunfeld and Griliches (1960), Hendry and Hubrich (2006), Kohn (1982), Lutkepohl (1984, 1987), and Pesaran, Pierse, and Kumar (1989). The theoretical literature provides somewhat inconclusive and often contradictory results. Under certain restrictions on the data generating processes (DGP) of the individual subcomponents, there might be gains in the forecasting ability of the AIM approach in terms of the mean square forecast error. This result is due to the larger information set used in the AIM approach. However, in general, the relative forecast efficiency of the approach will depend on the true DGPs. Under certain conditions, the forecasting efficiency of the AIM approach can actually be inferior to that of the aggregate forecast approach, due to the large number of parameters

estimated. As the cross-sectional dimension increases, so does the estimation uncertainty in the AIM approach, as the so-called "curse of dimensionality" makes it difficult to estimate the model accurately and raises the issue of the effect of estimation uncertainty on the model's forecasting performance. For recent surveys on the contemporaneous aggregation of disaggregate forecasts, see Granger (1990) and Lutkepohl (2006).

as the economic activity and geographical or locational characteristics of the specific region, and fall outside the control of airlines. *Service-related factors* are variables such as the quality and fare price components of air travel.

Our focus in this paper is on macroscopic models of air travel demand with geo-economic factors as explanatory variables. Macroscopic models attempt to model air travel demand in a particular region without considering the interactions between pairs of regions. These models can also be considered as aggregate models of air travel, and appear in the literature on the time series analysis of air travel demand more frequently than cross-sectional analysis. In most aggregate or macroscopic models, some measure of air travel demand, such as the total number of passengers or flights, or revenue per mile, is modeled as a linear function of various explanatory variables, including, but not limited to, some measure of price, some measure of an alternative mode of traveling, control variables such as the GDP of the region, some measure of tourism, some measure of foreign trade, etc. Among others, papers analyzing macroscopic models of air travel demand include Abed, Ba-Fail, and Jasimuddin (2001), Cline, Ruhl, Gosling, and Gillen (1998), Profillidis (2000), Saab and Zouein (2001) and Wang and Pitfield (1999). These models generally treat air travel demand as a homogenous commodity, and do not consider differences between air travel demand functions in different regions. In contrast to the previous literature, we model the ratio of the number of passengers originating from a given airport to the population served by that airport as a linear function of national and region-specific explanatory factors such as cost and economic activity.

The data used for this analysis were obtained from several different sources. The dependent variable, the ratio of the number of passengers originating from an airport to the population of the Metropolitan Statistical Area (MSA) served by that airport, has been calculated using data sets from the Bureau of Transportation Statistics and the US Census Bureau. In this section, we describe the sources of each variable in detail. We also discuss several data related issues.

In this paper, we use data for the 179 busiest airports in the US, which account for 97% of the total US air travel demand as of December 2004. This ratio is relatively constant over our sample period. To account for the residual air travel demand outside the areas served by these airports, we estimate the ratio of residual air travel demand to total air travel demand at the end of our in-sample period. We then assume that this ratio stays constant during the out-of-sample period. The detailed sources of each data set employed are discussed below.[3]

2.1. The dependent variable

The time series of the number of passengers originating from an airport is publicly available from the Bureau of Transportation Statistics at a monthly frequency. The database used in this analysis is the Air Carrier Statistics (Form 41 Traffic).[4] This database is frequently used by the aviation industry, the press, and the legislature to produce reports and analyses on air traffic patterns and carrier market shares, as well as passenger, freight and mail cargo flow within the aviation mode. A passenger is defined to be any person on board a flight who is not a member of the flight or cabin crew. The time series data are available from 1990 to 2004. The monthly time series for 179 airports and the total US air travel demand used in our analysis were filtered from this database.

We obtain the monthly population estimates for each MSA by aggregating population estimates of counties in a given MSA. We employ the MSA definitions given in the List of Metropolitan and Micropolitan Statistical Areas and Definitions, as of 2005.[5]

The annual county population estimates can be obtained from the US Census Bureau's Population Estimates Program archive. Each year, the Population Estimates Program produces total annual population estimates for each county. The annual county population estimates between the 1990 census in April 1990 and the 2000 census in April 2000 are based solely on the 1990 census, and do not reflect the 2000 census counts. Likewise, the annual county population

---

[3] The list of airports considered in this paper, along with details of the metropolitan statistical area served by each airport, is available from the authors upon request.

[4] T-100 Domestic Market (All Carriers) database of the Air Carrier Statistics (Form 41 Traffic) — All Carriers data library at the aviation web site of the Bureau of Transportation Statistics at http://www.transtats.bts.gov/.

[5] http://www.census.gov/population/www/estimates/metrodef.html.

estimates between the 2000 census and the end of our sample, December 2004, are all based on the 2000 census. To obtain monthly population estimates for each county between the 2000 census and the end of our sample, we linearly interpolate between the annual county population estimates from the Population Estimates Program archive. On the other hand, to obtain monthly population estimates between April 1990 and April 2000, we linearly interpolate between the 1990 and 2000 census counts, rather than linearly interpolating between the annual county population estimates. Although the annual county population estimates are also available to us, we choose to interpolate linearly between the two censuses due to discrepancies between the 1999 population estimates (which are based on the 1990 census) and the 2000 census population counts. For many counties in our data set, the annual population growth between 1999 and 2000, based on the 1999 population estimates and the 2000 census counts, was much higher than the average annual population growth between the 1990 and 2000 censuses, causing large jumps in our population estimates. To obtain monthly population estimates before April 1990, we linearly extrapolate based on the monthly population growth implied by the 1990 and 2000 censuses. Monthly population estimates for the United States are also available from the US Census Bureau's Population Estimates Program.

We use the number of passengers boarding commercial flights from a given airport in a given month as a proxy for air travel demand in the region served by that airport. We use the number of passengers rather than other possible measures of air travel demand that have been used in the literature. For example, Ito and Lee (2005) used revenue passenger miles (RPM) as a measure of aggregate air travel demand to analyze the effect of the September 11 terrorist attacks on US airline demand. Although we believe that RPM might provide a good proxy for aggregate air travel demand, it is not available for individual airports. Hence, we cannot employ RPM as a measure of air travel demand with our disaggregate approaches. Fig. 1 presents the monthly aggregate air travel demand in the US between January 1990 and December 2004.

Instead of modeling the number of passengers, we model the logit transformation of the ratio of monthly passenger and population estimates. The dependent variable can be considered as the daily propensity to travel or the per capita air travel demand in a given region.

Let pax_{it} denote the total number of passengers originating from airport i in month t, and pop_{it} be the month t population estimate of the MSA in which the airport is located, then the dependent variable is given by

$$y_{it} = \ln\left(\frac{(1/30)(\mathrm{pax}_{it}/\mathrm{pop}_{it})}{1 - (1/30)(\mathrm{pax}_{it}/\mathrm{pop}_{it})} \right) \qquad (2)$$

for $i = 1, 2, \ldots, 179$ and $t = 1, 2, \ldots, 180$ for the months between January 1990 and December 2004. This transformation helps to stabilize the variance, incorporates the implicit limitation on the number of trips that can be taken and facilitates the use of a model with the standard logistic-based origin-destination models. It is a fairly innocuous transformation, in the sense that predictions using per capita trips are similar and the main objective is forecasting a per capita rather than a total trip measure, since the latter may be driven by population growth.

2.2. Geo-economic factors

The geo-economic factors used as explanatory variables in this analysis can be grouped into three major categories depending on the geographical level of availability: the MSA, state and national levels.

2.2.1. MSA-level geo-economic factors

The unemployment rate for each MSA is obtained from the Bureau of Labor Statistics.[6] The Local Area Unemployment Statistics (LAUS) program at the BLS produces monthly and annual employment, unemployment, and labor force data for census regions, divisions, states, counties, metropolitan areas, and many cities, by place of residence. The monthly unemployment rate is obtained from LAUS. The percentage population growth is calculated at the MSA-level from the census population data described in Section 2.1.

2.2.2. State-level geo-economic factors

We employ the state-level coincident indicator index and the unemployment rate as a proxy for
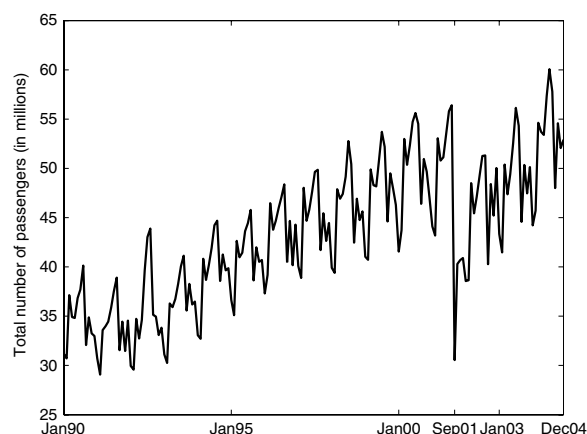
---

[6] http://www.bls.gov/lau/.

Fig. 1. Aggregate US air travel demand.
Note: The figure presents monthly numbers of passengers boarding commercial domestic flights in the US between January 1990 and December 2004.

the level of economic activity in a certain region. A coincident indicator summarizes several indicators such as nonagricultural employment, personal income and industrial production in a single index, thus indicating the current state of the economy in a certain region. We employ a Stock-Watson type[7] state-level coincident indicator index, which is available from the Federal Reserve Bank of Philadelphia at a monthly frequency.[8]

2.2.3. National-level geo-economic factors

We obtain the daily spot price of kerosene type jet fuel from the Energy Information Administration (EIA).[9] We obtain the jet fuel prices in cents per gallon as the average of the daily spot prices of Rotterdam (ARA) kerosene type jet fuel. We use Rotterdam spot prices rather than New York Harbor, US Gulf Coast or Los Angeles spot prices for data availability reasons. The Rotterdam spot prices have a correlation of 0.98 with the Los Angeles spot prices when both are available.

We obtain monthly prices (in dollars per barrel) of crude oil futures contracts from the Global Financial

---

[7] In the late 1980s, James Stock and Mark Watson developed a coincident index for the US economy as an alternative to the one which was published at that time by the Department of Commerce. The advantage of a Stock-Watson type index is that it combines several monthly indicators in a single measure of the economy.

[8] http://www.phil.frb.org/econ/stateindexes/index.html.

[9] http://www.eia.doe.gov/neic/historic/hpetroleum2.htm.

Data. We use the change in the average of the opening and closing prices in a given month as an explanatory variable in our empirical specification. We employ the futures contract to proxy for expected changes in future oil prices. These future prices may influence airline pricing of future flights, for which initial prices are typically set several months in advance.

3. Empirical specification and estimation

One can obtain monthly forecasts of the aggregate air travel demand, i.e., the total number of passengers boarding a commercial flight in the US in a given month, using either an aggregate approach or a disaggregate approach (when the disaggregate data are available). In this section, we describe the different empirical approaches employed for forecasting the aggregate air travel demand.

3.1. The empirical model

We assume that the air travel demand in a given region is a linear function of the regional geo-economic factors discussed above. More specifically, the empirical model is specified and estimated as a linear projection of the dependent variable y_{it} onto the lagged dependent variable and contemporary explanatory variables. Specifically, the model can be expressed as

$$y_{it} = \alpha_i + \beta_i' \mathbf{X}_{it} + \phi(L) y_{it} + \varepsilon_{it}, \qquad (3)$$

where y_{it} is the dependent variable (i.e., the propensity to travel in month t from airport i, as described in the data section), α_i is the constant term, $\mathbf{X}_{it}$ is a vector of explanatory variables, and $\phi(L)$ is a polynomial lag operator (i.e., $\phi(L) y_{it} = \sum_{j=1}^{l} \phi_{ij} y_{i,t-j}$). We estimate the same linear specification for the aggregate air travel demand data as for the disaggregate data. With the exception of the AIM approach, where we estimate the linear specification for each airport separately, the estimation of the linear specification for the disaggregate data is different to that for the aggregate data. It is these different estimation methods for the disaggregate data, which we discuss in further detail below, that allow us to exploit possible additional information in the disaggregate data.

We first estimate the contemporaneous model and iterate forward to obtain the h-step-ahead forecast.

Our forecasting approach differs from the usual textbook approach, due to the different estimation approaches used to estimate the linear specification with disaggregate data. We employ the iterative forecasting approach rather than the "h-step-ahead projection" approach implemented by Marcellino et al. (2003) because of the availability of forecasts for our explanatory variables. The advantage of the "h-step-ahead projection" with explanatory variables is that it eliminates the need to forecast the explanatory variables. In the iterative approach, one needs to forecast the explanatory variables in order to obtain forecasts of the dependent variable. In our analysis, one can obtain forecasts of the explanatory variables from other sources, such as government agencies. Furthermore, data availability becomes an issue for the "h-step-ahead projection" approach if one needs to forecast further into the future.

The implementation of Eq. (3) initially requires decisions to be made about which functions of the explanatory variables to employ and the order of the lag operator. Our modeling decisions are based on the goodness-of-fit of the model over the in-sample data. Specifically, the vector of explanatory variables, $\mathbf{X}_t$, includes a linear trend variable (t), the unemployment rate (unemp), the coincident indicator index (ci) and its square, the spot price of jet fuel (jetfuel), the monthly change in the average price of crude oil futures (Δ oilfutures), and monthly dummy variables (sd$_k$, where $k = 1, 2, \ldots, 11$ for monthly seasonal dummies) to account for the seasonal pattern of air travel demand. To account for the effect of September 11 on air travel demand in the US, we follow Ord and Young (2004) and model the effect as a temporary change, which assumes that the factor has a relatively short-term impact on the series. Specifically, we assume that the September 11 attacks affected only the last two-thirds of September and the whole month of October. In other words, we set the September 11 effect to 2/3 for the month of September and 1 for the month of October,[10] and assume that the effect decreases exponentially thereafter. More specifically, we model the effect of September 11 on air travel

demand as follows:

$$\text{sept11}_t = \begin{cases} 0, & \text{if } t < \tau, \\ 2/3, & \text{if } t = \tau, \\ d^{t-\tau-1}, & \text{if } t > \tau, \end{cases} \quad (4)$$

where τ is the event period, i.e., September 2001, and d is the adjustment factor, chosen to be 0.6.[11] We choose the number of lags in our specification to be 1, in order to account for first order autocorrelation in air travel demand. The empirical model estimated can be expressed as follows[12]:

$$\begin{aligned} y_{it} = {} & \alpha_i + \beta_{i1}t + \beta_{i2}\text{unemp}_{it} + \beta_{i3}\text{ci}_{it} + \beta_{i4}\text{ci}_{it}^2 \\ & + \beta_{i5}\text{jetfuel}_t + \beta_{i6}\Delta\text{oilfutures}_t + \beta_{i7}\text{sept11}_t \\ & + \sum_{k=1}^{11} \theta_{ik}\text{sd}_k + \phi_{i1}y_{i,t-1} + \varepsilon_{it}. \end{aligned} \quad (5)$$

The intuition behind this empirical specification is simple. The trend term controls for an increase over time in the per capita propensity to travel. Both the unemployment rate and the coincident indicator are used to account for the effect of economic activity on per capita air travel demand. The spot price of jet fuel can be considered as a cost factor to passengers and airlines. However, it can also be considered as an indicator of the overall economic activity, since it is highly correlated with crude oil prices. The oil futures price is used to account for possible changes in future oil prices. Most airlines plan their flight schedules and airfares three months or more in advance, and thus an increase in futures prices on crude oil should imply an upward pressure on average airfares as airlines try to pass on their increased costs in the form of fare increases.

3.2. Estimating the empirical model

In order to forecast the aggregate US air travel demand, we estimate the empirical model for aggregate and disaggregate data. We first split our sample into two subsamples: an in-sample period

---

[10] We also experimented with a value of 1 for the month of September rather than 2/3, but our results did not change in any significant fashion.

[11] We also experimented with other values for the adjustment factor, but our results did not change qualitatively.

[12] Variables with a subscript of i indicate that they are region specific. Note however that the coincident indicator (ci) available at the state level is the same for all airports in the same state, even though the variable has a subscript of i.

for estimating the coefficients of our empirical model and an out-of-sample period for evaluating the different forecasts of aggregate air travel demand. The in-sample period is between 1990 and 2002, resulting in 156 monthly observations for each airport and for the aggregate air travel demand. In the benchmark study, the out-of-sample period is between 2003 and 2004, resulting in a total of 24 monthly observations for evaluating our out-of-sample forecasts. For consistency, we decided to always use the same in-sample period, and instead change the out-of-sample period for different forecast horizons and approaches.

In this section, we discuss the different methods used to estimate the coefficients from both the aggregate data and the individual airport data. The forecasting approaches differ only with respect to the estimation method used. Once the coefficients have been estimated, the forecasting step is identical across different approaches.

3.2.1. The aggregate approach

The empirical model in Eq. (5) for the aggregate US air travel demand is estimated using aggregate explanatory variables via ordinary least squares. Table 1 summarizes the estimation results for the aggregate US air travel demand for the in-sample period.

Several interesting facts emerge from the aggregate estimation results. First of all, our empirical specification captures 94% of the variation in per capita air travel demand, suggesting a good in-sample fit. Although it is insignificant, there is a negative linear trend in the aggregate per capita air travel demand when we control for the autoregressive terms in the specification. As expected, the level of economic activity, as measured by the national coincident indicator, has a positive first degree effect on the per capita air travel demand. Not surprisingly, the per capita air travel demand was negatively affected by the September 11 shock. Although the estimates are not significant at any conventional level, the cost-related factors (jet fuel spot prices and oil futures prices) have positive effects on the air travel demand. This result might be due to the positive correlation between cost-related factors and the level of economic activity. As expected, the air travel demand is higher during the summer months, as is indicated by positive coefficient estimates on the dummy variables for the summer months.

3.2.2. The AIM approach

In the AIM approach, we assume that each individual airport has unique dynamics and that the empirical model is estimated separately for each airport. In other words, this approach assumes that individual airports do not have common factors and that the effects of each explanatory variable, whether national or airport-specific, are different across airports.

In all disaggregate approaches, including the AIM approach, if an explanatory variable such as the unemployment rate is available at the MSA level, we use MSA-level data. This is generally airport specific, except for MSAs with more than one airport, such as the New York-Northern New Jersey-Long Island MSA. If MSA-level data are not available for an explanatory variable, we use data for the state in which the airport is located, such as the coincident indicator. Finally, if the variable is not available at either MSA or state level, we use national-level data for that variable, for example jet fuel and oil futures prices. These variables are identical for every airport and the exact same data are used in the aggregate approach. An index of i on a variable in Eq. (5) implies that the variable is available at either the MSA or state level. Otherwise, the variable is only available at the national level.

The advantage of the AIM approach relative to other approaches is that it allows for heterogeneity across different airports while still using individual market level data to forecast the aggregate air travel demand. However, one should note that this might not be the most efficient way of employing the heterogeneity, due to the relatively large number of coefficients estimated. We need to estimate the empirical model for each individual airport. Hence, for monthly data we need to estimate 20 coefficients for each airport using only 155 observations, after adjusting for lags. The ratio of the number of coefficients estimated to the number of observations in the AIM approach is identical to that of the aggregate approach.

Table 2 presents summary statistics for the coefficient estimates from estimating the empirical specification for individual air travel demands separately. The coefficient estimates from the AIM approach allow us to identify possible sources of heterogeneity in air travel demands across different airports. Based on the standard deviation of the estimated coefficients for individual airports, the reaction to September 11,
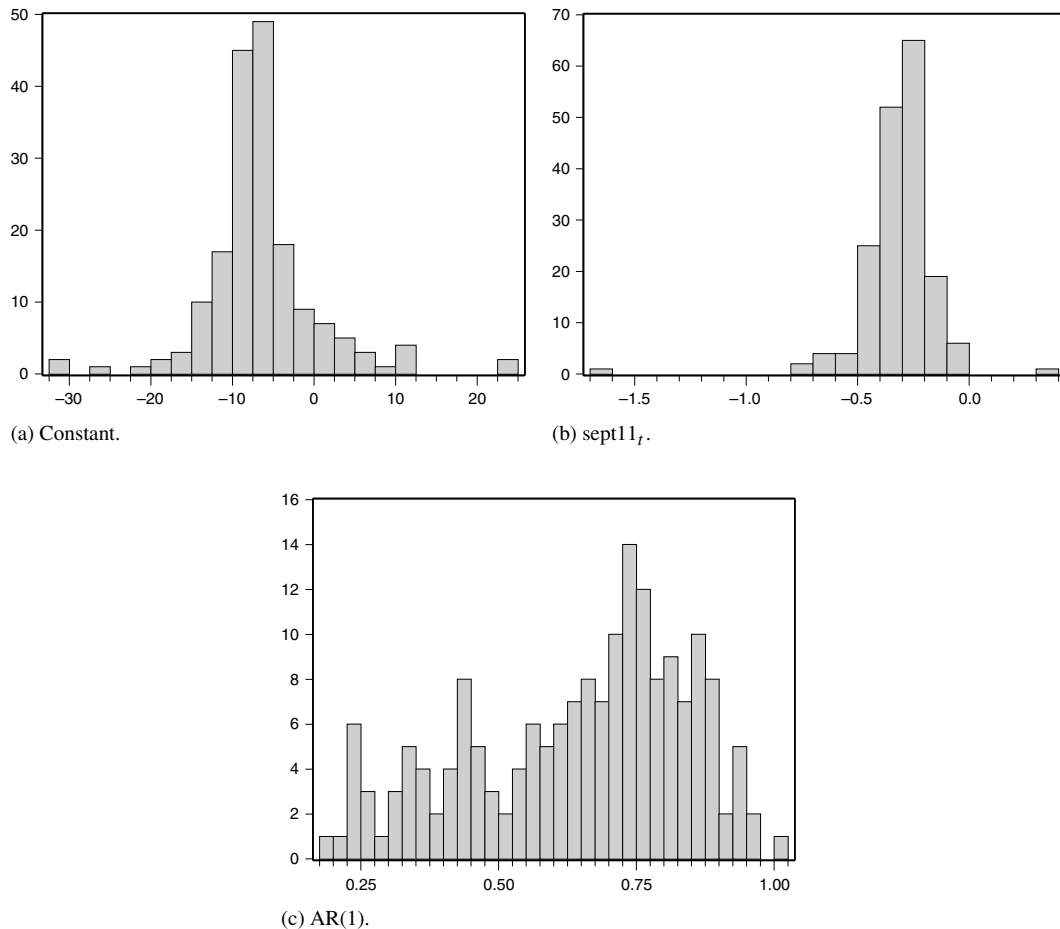
Fig. 2. Histograms of selected coefficient estimates from the AIM approach.

Note: The figure presents histograms of selected coefficient estimates from Eq. (5). The empirical model in Eq. (5) is estimated separately for individual airports using disaggregate level data.

the autocorrelation, the effects of the summer months, and the constant term might all be possible sources of heterogeneity across different airports. This can also be seen easily from the histograms of estimated coefficients. Fig. 2 presents the histograms of selected coefficient estimates for individual airports. Although the AIM approach may not be the most efficient way to exploit this heterogeneity due to estimation uncertainty, it should provide some intuition as to the possible sources of the heterogeneity and efficient ways of exploiting it.

3.2.3. The quasi-AIM approach

In this section, we discuss several estimation approaches which we collectively term "quasi-AIM".

Unlike the AIM approach, which attempts to fully exploit the heterogeneity across individual airports, the quasi-AIM approaches exploit the heterogeneity partially, by restricting the effects of certain variables to be identical across individual airports, in order to avoid the problems associated with estimation uncertainty. The quasi-AIM approaches differ with respect to the degree and source of the heterogeneity used in the estimation of the empirical model.

In all quasi-AIM approaches, the empirical model is estimated via pooled least squares over the in-sample period using the panel of individual airport data. The empirical model is estimated by restricting all coefficient estimates to be the same across different airports, except for those that are allowed to be

Table 1

Coefficient estimates from the aggregate approach.

Variable	Coefficient	t-statistic	Prob.
α	−10.4881	−5.53	0.0000
t	−0.0043	−1.45	0.1488
$unemp_t$	0.0170	0.58	0.5639
ci_t	0.0735	3.08	0.0025
ci_t^2	−0.0002	−3.80	0.0002
$jetfuel_t$	0.0004	1.58	0.1175
$\Delta\,oilfutures_t$	0.0007	0.45	0.6553
$sept11_t$	−0.3036	−4.47	0.0000
January dummy	−0.1199	−3.85	0.0002
February dummy	−0.1310	−4.76	0.0000
March dummy	0.0631	2.79	0.0061
April dummy	0.0246	1.80	0.0739
May dummy	0.0442	3.75	0.0003
June dummy	0.0868	5.11	0.0000
July dummy	0.1275	8.01	0.0000
August dummy	0.1378	9.30	0.0000
September dummy	−0.0672	−2.94	0.0039
October dummy	0.0441	2.91	0.0042
November dummy	−0.0174	−1.19	0.2347
AR(1)	0.2585	1.24	0.2163
R^2	0.9385	F-statistic	108.4000
Adjusted R^2	0.9298	Residual SE	0.0324

Note: The table presents the coefficient estimates of Eq. (5) from the aggregate approach using aggregate per capita air travel demand and aggregate explanatory variables for the in-sample period between January 1990 and December 2002. "Coefficient" and "t-statistic" correspond to the estimate and the t-statistic, respectively. "Prob." is the p-value associated with a two-sided test based on the t-statistic, and "Residual SE" is the standard deviation of the residuals.

airport specific. This approach allows us to exploit the information available across different airports to a certain extent, without suffering from as much estimation uncertainty as the AIM approach. The number of coefficients estimated is either 198 or 376, depending on the number of unrestricted coefficients in the corresponding quasi-AIM model.

Based on the results from the AIM estimation, we consider the following quasi-AIM approaches:

1. Pooled: Disaggregated model with common coefficients across all airports;
2. Quasi-AIM (FE): Disaggregated model with common coefficients across all airports except for the constant term (fixed effect);
3. Quasi-AIM (sept11): Disaggregated model with common coefficients across all airports except for the coefficient on the September 11 variable;
4. Quasi-AIM (AR(1)): Disaggregated model with common coefficients across all airports except for the coefficient on the lagged dependent variable;
5. Quasi-AIM (FE & sept11): Disaggregated model with common coefficients across all airports except for the constant term (fixed effect) and the coefficient on the September 11 variable;
6. Quasi-AIM (FE & AR(1)): Disaggregated model with common coefficients across all airports except for the constant term (fixed effect) and the coefficient on the lagged dependent variable; and
7. Quasi-AIM (sept11 & AR(1)): Disaggregated model with common coefficients across all airports except for the coefficients on the September 11 and lagged dependent variables.

Different quasi-AIM approaches exploit the heterogeneity across airports in different dimensions. For example, the quasi-AIM sept11 approach exploits the possible heterogeneity between airports with respect to their reaction to the September 11 attacks. One might argue that airports with mostly domestic and particularly short-haul commuter traffic would have been affected the most by September 11, whereas airports with mostly international traffic would have been relatively less affected. The quasi-AIM AR(1) approach exploits the heterogenous information available in the autoregressive dynamics of different airports. Consider two airports, say Las Vegas and San Francisco, where Las Vegas is a growing market and San Francisco is a mature one with respect to per capita demand. Although these two individual airports might be affected similarly by explanatory variables such as the economic conditions and jet fuel prices, it is still reasonable to assume that they will have different autoregressive dynamics. We also analyze three additional quasi-AIM approaches where we allow two of these three factors to be airport-specific while still restricting the coefficient estimates of all other variables to be identical. In other words, we attempt to exploit the heterogeneity in two dimensions at the same time, rather than in a single dimension.

Finally, we also analyze the most restrictive form of the quasi-AIM approach, which we term the pooled approach. It uses all available data at the individual level but imposes the restriction that all coefficient estimates are identical across airports. In the pooled approach, as the name suggests, we estimate the empirical model via pooled least squares

Table 2
Summary statistics for coefficient estimates from the AIM approach.

Variable	Mean	Median	Min	Max	Std. dev.
α	−6.4534	−7.0302	−31.9936	24.5534	7.0399
t	−0.0004	−0.0008	−0.0603	0.0246	0.0075
$unemp_t$	−0.0054	−0.0059	−0.0706	0.0628	0.0149
ci_t	0.0078	0.0192	−0.5727	0.5502	0.1268
ci_t^2	0.0000	−0.0001	−0.0028	0.0028	0.0006
$jetfuel_t$	−0.0001	0.0000	−0.0050	0.0035	0.0011
$\Delta\,oilfutures_t$	0.0014	0.0013	−0.0171	0.0202	0.0039
$sept11_t$	−0.3148	−0.2915	−1.6726	0.3242	0.1614
January dummy	−0.1159	−0.1328	−0.3011	0.6218	0.1175
February dummy	−0.1226	−0.1432	−0.4014	0.6156	0.1344
March dummy	0.0567	0.0477	−0.2230	0.7773	0.1366
April dummy	−0.0049	0.0000	−0.8477	0.6720	0.1456
May dummy	0.0311	0.0441	−1.0598	0.6816	0.1384
June dummy	0.0943	0.0992	−0.7540	0.6623	0.1315
July dummy	0.1396	0.1291	−1.0846	0.8805	0.1849
August dummy	0.1420	0.1165	−1.0996	1.0168	0.2051
September dummy	−0.0421	−0.0513	−0.8347	0.7330	0.1546
October dummy	0.0641	0.0606	−0.3356	0.5647	0.1187
November dummy	−0.0182	−0.0099	−1.0764	0.2659	0.1074
AR(1)	0.6465	0.6960	0.1915	1.0128	0.1990
Residual SE	0.0706	0.0604	0.0356	0.4039	0.0413

Note: The table presents summary statistics on the coefficient estimates of Eq. (5) from the AIM approach using the per capita air travel demand and explanatory variables for the individual airports between January 1990 and December 2002. "Std. dev." is the standard deviation across airports of the coefficient estimate of the variable in the corresponding row. "Residual SE" is the standard deviation of residuals.

using individual level dependent and explanatory variables. In other words, the empirical model in Eq. (5) is estimated separately for individual airport data by restricting the coefficients to be identical across airports. The advantage of this approach with respect to the aggregate approach is the availability of a larger data set. The number of coefficients to be estimated is 20, whereas the number of observations in the in-sample data set, after adjusting for autoregressive terms, is 27,745.[13] Although the pooled approach employs all available data, it does not take into account the information embedded in the heterogeneity across airports. In other words, this approach does not efficiently use all available information. Table 3 summarizes the estimation results for the pooled estimation of the empirical model.

The coefficient estimates obtained by the pooled approach are somewhat different to those obtained by the aggregate approach. In the pooled estimation, the effect of economic activity on per capita air travel demand is captured by the unemployment rate, which has a significant negative impact. On the other hand, the other measure of economic activity, namely the state-level coincident indicator, has an insignificant effect on air travel demand, in contrast to its significant effect on the aggregate air travel demand. This might be due to the fact that the unemployment rate is available at the MSA level, and therefore might be a better measure of economic activity for an individual airport than the coincident indicator, which is only available at the state level. On the other hand, variables such as the change in oil futures prices and the lagged per capita air travel demand become significant. The R^2 of the pooled estimation is 0.98, suggesting an almost perfect in-sample fit. Most of the coefficient estimates that are insignificant using aggregate data become significant in the pooled estimation. These differences show the effect on coefficient estimates of using disaggregate data.

Between the two extremes of the pooled and AIM approaches, one might argue that the quasi-AIM approaches provide a better way of exploiting

---

[13] The number of time series observations available for each airport after adjusting for lags ($156 − 1 = 155$), times the number of airports in our sample (179).

Table 3
Coefficient estimates from the pooled approach.

Variable	Coefficient	t-statistic	Prob.
α	−6.3075	−6.34	0.0000
t	0.0036	1.45	0.1467
unemp_t	−0.0034	−1.98	0.0479
ci_t	−0.0060	−0.47	0.6419
ci_t^2	0.0000	0.84	0.3985
jetfuel_t	−0.0002	−1.25	0.2123
Δ oilfutures$_t$	0.0017	2.66	0.0077
sept11_t	−0.3050	−15.77	0.0000
January dummy	−0.1134	−39.60	0.0000
February dummy	−0.1201	−35.74	0.0000
March dummy	0.0588	15.02	0.0000
April dummy	−0.0028	−0.57	0.5663
May dummy	0.0333	6.67	0.0000
June dummy	0.0965	18.47	0.0000
July dummy	0.1416	27.88	0.0000
August dummy	0.1440	29.41	0.0000
September dummy	−0.0407	−8.64	0.0000
October dummy	0.0654	17.80	0.0000
November dummy	−0.0175	−6.15	0.0000
AR(1)	0.9907	661.99	0.0000
R^2	0.9832	F-statistic	85275
Adjusted R^2	0.9832	Residual SE	0.1411

Note: The table presents the coefficient estimates of Eq. (5) from the pooled approach using the per capita air travel demand and explanatory variables for the individual airports between January 1990 and December 2002. See also the notes to Table 1.

the heterogeneity without being subject to possible estimation uncertainty. The quasi-AIM approach allows for a certain degree of heterogeneity which is not possible with the pooled approach, and it is not subject to as much estimation uncertainty as the AIM approach, since the number of coefficients to be estimated is significantly smaller. In other words, the quasi-AIM exploits several dimensions of possible heterogeneity across airports without being subject to extreme estimation uncertainty. One can use the estimation results from the pooled and AIM approaches to discover the possible sources of most of the heterogeneity across individual markets.

3.3. Benchmark models

We also obtain out-of-sample forecasts of the aggregate air travel demand from benchmark models without any explanatory variables except for the trend and the September 11 variable. In doing this, we analyze whether the explanatory variables provide any out-of-sample predictive power in our setting.

Furthermore, it is fairly common in the forecasting literature for parsimonious forecasting approaches to outperform more complicated approaches in terms of their out-of-sample forecasting performances. The benchmark models are based on an empirical specification which is identical to the original empirical specification in Eq. (5), except for the explanatory variables:

$$y_{it} = \alpha_i + \beta_{i1}t + \beta_{i2}\text{sept}11_t + \sum_{k=1}^{11} \theta_{ik}\text{sd}_k$$
$$+ \phi_{i1}y_{i,t-1} + \varepsilon_{it}. \tag{6}$$

Benchmark models also differ from each other as to which of the estimation approaches discussed above is used. For the sake of brevity, we do not present estimation results for the benchmark models.

4. Forecasting the aggregate air travel demand

In order to form out-of-sample forecasts of the aggregate air travel demand in the US, we need forecasts of the independent variables. In this section, we discuss the way in which out-of-sample forecasts of the independent variables are obtained and forecasts of the aggregate air travel demand are obtained from the different approaches discussed above.

4.1. Forecasting independent variables

For most of the explanatory variables considered in this paper, we can employ forecasts from other sources such as government agencies. In this section, we discuss the sources of the out-of-sample forecasts of the explanatory variables and the assumptions underlying them.

We assume that the unemployment rate for individual MSAs is constant, and is equal to the unemployment rate in the last period of the corresponding in-sample data. The MSA level unemployment rate is quite persistent, which makes our assumption of a constant MSA level unemployment reasonable. Following the Congressional Budget Office's forecast of a 3% annual growth rate for the US real output between 2003 and 2004, we assume that the state-level coincident indicator indices grow at a monthly rate of 0.2466%. This forecast of a 3% annual growth rate is also consistent with the forecasts of other government agencies such as the Census Bureau, the FAA and the

Administration, as well as with the Blue Chip forecasts. Our monthly forecasts of jet fuel prices and oil futures prices are based on the Energy Information Agency's (EIA) Annual Energy Outlook (2003).[14] In the reference case, the EIA forecast that the world oil price would increase at an annual rate of 1.43%, and a corresponding monthly rate of 0.1184%. We assume that jet fuel prices and oil futures prices are perfectly correlated with the world oil price and grow at the same rate. This assumption is reasonable, since these variables are highly correlated with the world oil price in our sample. Our population forecasts for MSAs are based on the Census Bureau's MSA-level forecasts. The Census Bureau employs the 2000 census to obtain population projections.

Ex-post, these forecasts generally underpredicted the corresponding explanatory variables. Nevertheless, these forecasts provide a sensible benchmark. One can easily extend this analysis by analyzing different scenarios such as the low vs. high world price of oil and the low vs. high growth of the economy. Furthermore, in a model with a relatively large number of time series observations and a small number of autoregressive terms, one can consider employing the "*h*-step-ahead projection" of Marcellino et al. (2003). We should note that the choice of forecasting approach for the explanatory variables does not seem to have a significant effect on our conclusions about the relative performances of different approaches in forecasting the aggregate air travel demand.

4.2. Forecasting the aggregate air travel demand

Forecasting the aggregate air travel demand requires forecasts of the per capita air travel demand at either the aggregate or disaggregate level, as well as population forecasts. We model the per capita air travel demand as the dependent variable in our empirical specifications instead of the air travel demand itself. Hence, we first form out-of-sample forecasts of the per capita air travel demand via either the aggregate or disaggregate approach, then convert the forecasts

of the per capita air travel demand into forecasts of numbers of passengers by using out-of-sample population forecasts and the inverse of the logit transformation in Eq. (2).

Forecasting the aggregate air travel demand via the aggregate approach is relatively straightforward. We just convert the forecasts of the aggregate per capita air travel demand into forecasts of the number of passengers, as discussed above. On the other hand, forecasting the aggregate air travel demand via one of the disaggregate approaches requires forecasts of the air travel demands at individual airports. The aggregate air travel demand is, by definition, the sum of the individual air travel demands at all US airports. However, we only forecast the air travel demands at the 179 busiest airports, rather than at all US airports. Hence, forecasts of the aggregate air travel demand are obtained as the sum of the air travel demands of the 179 busiest airports, adjusted by a factor that accounts for the air travel demand at the other airports. As was discussed in Section 2, this adjustment factor is the reciprocal of the ratio of the air travel demands at the 179 busiest airports to the total US air travel demand in the last month of the in-sample period, and is assumed to be constant in the out-of-sample periods.

As an initial exercise, we analyze the forecasting performances of the different approaches based on their Mean Absolute Forecast Errors (MAFE) and Root Mean Square Errors (RMSFE) for the out-of-sample period between January 2003 and December 2004. To do so, we first estimate the empirical specification using the in-sample data between January 1990 and December 2002. Assuming that there are no structural breaks in the out-of-sample period, we then form one-step-ahead forecasts of the per capita air travel demand and proceed in an iterative fashion to obtain *h*-step-ahead forecasts which depend on the $h-1$-step-ahead forecasts and parameter estimates based on the in-sample period between January 1990 and December 2002. The first exercise can be thought of as a static forecasting exercise, where the parameter estimates are based on a fixed window, are not updated and do not depend on the forecast horizon. Specifically, the forecasting model can be expressed as follows:

$$\hat{y}_{i,t+h} = \hat{\alpha}_i + \hat{\beta}_{i1}(t+h) + \hat{\beta}_{i2}\widehat{unemp}_{i,t+h}$$
$$+ \hat{\beta}_{i3}\widehat{ci}_{i,t+h} + \hat{\beta}_{i4}\widehat{ci}^2_{i,t+h} + \hat{\beta}_{i5}\widehat{jetfuel}_{t+h}$$

---

[14] The Annual Energy Outlook presents midterm forecasts of energy supply, demand, and prices through until 2025, prepared by the EIA. The projections are based on results from the EIA's National Energy Modeling System (NEMS). The forecast employs only the data available at the end of our in-sample period, since the Annual Energy Outlook 2003 was published in January 2003 (see http://tonto.eia.doe.gov/FTPROOT/forecasting/0383(2003).pdf).

$$+ \hat{\beta}_{i6} \Delta \widehat{\text{oilfutures}}_{t+h} + \hat{\beta}_{i7} \widehat{\text{sept11}}_{t+h}$$

$$+ \sum_{k=1}^{11} \hat{\theta}_{ik} \text{sd}_k + \hat{\phi}_{i1} \hat{y}_{i,t+h-1}, \tag{7}$$

where the parameter estimates are based on the fixed in-sample period between January 1990 and December 2002. h-step-ahead out-of-sample forecasts of the explanatory variables can be obtained as discussed in Section 4.1, and are denoted as for the unemployment rate for airport i, which is given by $\widehat{\text{unemp}}_{i,t+h}$.

Although the first forecasting exercise might provide an indication as to which model is better, one should be careful in interpreting the MAFEs and RMSFEs from the benchmark case, as they are calculated by averaging out the forecast errors for different forecast horizons. Furthermore, in a recent paper, Jordà and Marcellino (2010) questioned the validity of using these loss functions for a path forecast, as in our first exercise. Finally, the first exercise does not reveal much information about the relative forecasting performances of different approaches for a certain forecast horizon. In our second forecasting exercise, we analyze the out-of-sample forecasting performances of different approaches for three different forecast horizons, namely short-, medium- and long-term forecast horizons, which correspond to 1-, 6- and 12-step-ahead forecasts, respectively. The second approach allows us to analyze the forecasting performance for certain forecast horizons, and by doing so, one might gain additional insights as to which model should be used for a given forecast horizon.

In the second forecasting exercise, we form forecasts of the per capita air travel demand in a recursive fashion, using expanding windows of in-sample data. Whether we are producing 1-, 6- or 12-step-ahead forecasts, for consistency reasons, we use data between January 1990 and December 2002 as the initial window in a series of expanding windows of in-sample observations. The forecasting approach is similar for short-, mid- and long-term forecasts, and can be described briefly for 6-step-ahead forecasts as follows. In order to form forecasts of the air travel demand in June 2003, we first estimate the coefficients using the in-sample data between January 1990 and December 2002. We then forecast the air travel demand as in the first forecasting exercise.

For July 2003, we estimate the coefficients using data between January 1990 and January 2003. In other words, we expand the in-sample data by adding observations from January 2003, and re-estimate the coefficients every time we expand the in-sample data. We continue in this fashion until December 2004, the last month of our out-of-sample period. The second forecasting exercise can be considered as a dynamic forecasting approach where the parameter estimates are updated as new observations arrive.

5. Forecasting results

In this section, we discuss the out-of-sample forecasting performances of the different approaches from the different forecasting exercises. The out-of-sample forecasting performance of each approach depends on the tradeoff between estimation uncertainty and the degree of heterogeneity allowed. The aggregate and the pooled approaches do not allow for any heterogeneity across airports in the estimation step. The estimation uncertainty of these approaches is relatively low, since the number of coefficients estimated is low relative to the quasi-AIM and AIM approaches. On the other hand, the AIM approach employs all available heterogeneity by estimating the empirical model without restrictions. Hence, the estimation uncertainty is arguably high relative to other approaches. The quasi-AIM approaches can be considered as being a middle ground between the pooled and AIM approaches. The quasi-AIM approaches partially exploit the heterogeneity across airports, without having to estimate a relatively large number of coefficients. In one sense, the quasi-AIM approaches can be considered as exploiting the heterogeneity across airports more efficiently without suffering from too much estimation uncertainty.

We first present the results from the first (static) forecasting exercise discussed above, where we do not distinguish between the different forecast horizons. We use the mean absolute forecast error (MAFE) and the root mean square forecast error (RMSFE) as measures of the out-of-sample forecasting performance. The mean absolute forecast error is defined as the average of the absolute nominal forecast errors and the root mean square forecast error is defined as the square root of the average of the square nominal forecast errors, where the nominal forecast error is the difference between the actual

number of passengers in a given month and the forecast number of passengers in that month. As was discussed in more detail above, one should be careful in interpreting measures of forecasting performances based on the MAFE and RMSFE for different forecast horizons. Table 4 summarizes the relative forecasting performances of the different approaches with respect to the aggregate approach.

The quasi-AIM approach with fixed effects and different reactions to the September 11 terrorist attacks performs best in terms of out-of-sample forecasting ability, as measured by the mean absolute forecast error. The quasi-AIM approaches which exploit the heterogeneity in these two dimensions separately, i.e., the quasi-AIM approach with fixed effects only and the quasi-AIM approach with different September 11 coefficients, have similarly good out-of-sample performances. These results are consistent with our in-sample findings based on the AIM approach, where we identify the reaction to September 11 and the level of per capita air travel demand (fixed effects) among the possible sources of heterogeneity across different airports. On the other hand, any quasi-AIM approaches that attempt to exploit the heterogeneity in the autocorrelation structure of the per capita travel across different airports perform relatively poorly and provide worse static out-of-sample forecasts than the pooled and AIM approaches. Furthermore, the pooled approach provides slightly better out-of-sample forecasts than the AIM approach in terms of MAFE. This result suggests that the pooled approach might be a better way of employing the information embedded in disaggregate level data than the AIM approach, which may suffer from high estimation uncertainty.

More importantly, all approaches that employ information which is available at the disaggregate level perform better than the aggregate approach. In addition, all approaches perform better than their corresponding benchmark approach without any explanatory variables, except for the aggregate approach and any quasi-AIM approach with different AR(1) coefficients. These results suggest that using explanatory variables in the aggregate approach has a negative effect on the out-of-sample forecasting performance of that approach. In other words, one might obtain better out-of-sample forecasts of the aggregate variable by using a simple empirical approach based on autore-

Table 4
Out-of-sample performance of static forecasts.

(a) Relative MAFEs and RMSFEs with respect to the aggregate approach.

	MAFE	RMSFE	Rank
Aggregate	1.0000	1.0000	9
Pooled	0.4351	0.4178	4
Quasi-AIM (FE)	0.4247	0.4147	2
Quasi-AIM (sept11)	0.4319	0.4145	3
Quasi-AIM (AR(1))	0.5878	0.5747	7
Quasi-AIM (FE & sept11)	0.4240	0.4137	1
Quasi-AIM (FE & AR(1))	0.5970	0.6051	8
Quasi-AIM (sept11 & AR(1))	0.5492	0.5286	6
AIM	0.4960	0.5122	5

(b) Relative MAFEs and RMSFEs with respect to the corresponding benchmark model.

	MAFE	RMSFE
Aggregate	1.3509	1.4449
Pooled	0.9157	0.9152
Quasi-AIM (FE)	0.8228	0.7961
Quasi-AIM (sept11)	0.9140	0.9125
Quasi-AIM (AR(1))	1.1319	1.1679
Quasi-AIM (FE & sept11)	0.8114	0.7874
Quasi-AIM (FE & AR(1))	1.3461	1.4137
Quasi-AIM (sept11 & AR(1))	1.0740	1.0899
AIM	0.5523	0.6392

Note: The table presents the out-of-sample performances of static forecasts for the aggregate air travel demand. The static forecasts are formed in a recursive fashion based on the in-sample data between January 1990 and December 2002, and the coefficient estimates are not updated. Panel (a) presents the relative performances of the static forecasts listed in rows, with respect to the aggregate approach. The MAFE and RMSFE columns report the relative mean absolute forecast errors and the root mean square forecast errors, respectively. The relative measures are obtained by dividing the MAFE or RMSFE of the forecasting approach by those of the aggregate approach. A forecasting approach outperforms the aggregate approach if and only if its relative MAFE or RMSFE is less than one. "Rank" is the rank based on the MAFE of each forecasting approach relative to other approaches, where 1 corresponds to the best performing approach, i.e., the smallest relative MAFE. Panel (b) presents the relative MAFE and RMSFE of each forecasting approach with respect to the corresponding benchmark model without explanatory variables which need to be forecast. The relative measures are obtained by dividing the MAFE or RMSFE of the forecasting approach by those of the corresponding benchmark model. A number less than 1 suggests that the explanatory variables provide some out-of-sample forecasting power.

gressive dynamics, rather than an empirical approach where the explanatory variables need to be forecast. Furthermore, based on the out-of-sample forecasting performance, using explanatory variables in our

Table 5

Out-of-sample performances of dynamic short-, medium- and long-term forecasts.

	Short-term (1-step-ahead)			Medium-term (6-step-ahead)			Long-term (12-step-ahead)		
	MAFE	RMSFE	Rank	MAFE	RMSFE	Rank	MAFE	RMSFE	Rank
Aggregate	1.0000	1.0000	9	1.0000	1.0000	9	1.0000	1.0000	9
Pooled	0.5311	0.5602	7	0.3979	0.4480	7	0.2575	0.3170	2
Quasi-AIM (FE)	0.4387	0.4531	2	0.3732	0.4219	3	0.2745	0.3182	6
Quasi-AIM (sept11)	0.5335	0.5631	8	0.3978	0.4477	6	0.2572	0.3167	1
Quasi-AIM (AR(1))	0.4602	0.4751	4	0.3971	0.4446	5	0.2678	0.3172	4
Quasi-AIM (FE & sept11)	0.4392	0.4535	3	0.3730	0.4217	2	0.2740	0.3178	5
Quasi-AIM (FE & AR(1))	0.4287	0.4440	1	0.3678	0.4205	1	0.2817	0.3242	7
Quasi-AIM (sept11 & AR(1))	0.4710	0.4942	5	0.3967	0.4432	4	0.2624	0.3145	3
AIM	0.4989	0.5094	6	0.5079	0.5497	8	0.4798	0.5215	8

Note: The table presents the out-of-sample performances of dynamic forecasts for the aggregate air travel demand. The dynamic forecasts are formed in a recursive fashion based on expanding windows of in-sample observations. The short forecast horizon corresponds to 1-step-ahead forecasts; the medium forecast horizon corresponds to 6-step-ahead forecasts; and the long forecast horizon corresponds to 12-step-ahead forecasts. The MAFE and RMSFE columns report the relative mean absolute forecast errors and the root mean square forecast errors, respectively. The relative measures are obtained by dividing the MAFE or RMSFE of the forecasting approach by that of the aggregate approach. A forecasting approach outperforms the aggregate approach if and only if its relative MAFE or RMSFE is less than one. "Rank" is the rank based on the MAFE of each forecasting approach relative to other approaches, where 1 corresponds to the best performing approach, i.e., the smallest relative MAFE.

empirical specification improves the out-of-sample forecasting power at the disaggregate level for most approaches. The explanatory variables provide significant additional out-of-sample forecasting power over the corresponding benchmark in forecasting the aggregate air travel demand, although we did need to forecast these explanatory variables, which might have had a negative effect on the forecasting power ex ante. These results might also justify these explanatory variables and their forecasts provided by government agencies in forecasting aggregate air travel demand.

In the second (dynamic) exercise, we analyze the forecasting performances of different approaches for three different forecast horizons, short- (1-step-ahead), medium- (6-step-ahead) and long-term (12-step-ahead). This exercise allows us to identify not only the best forecasting approach for different forecast horizons, but also possible reasons for the relative performances in the first exercise. To this end, Table 5 presents the out-of-sample forecasting performances in terms of relative MAFEs and RMSFEs with respect to the aggregate approach, as well as presenting the relative rank of each approach for a given forecast horizon.

The quasi-AIM approach with different autocorrelation coefficients across airports (Quasi-AIM (AR(1))) is the best forecasting approach for short- and medium-term forecasting. On the other hand, it is among the worst performers for long-term forecasts,

a fact which might explain the relatively poor performance of this approach in the first forecasting exercise, where we did not distinguish between the different forecast horizons. The good performance of this approach for short- and medium-term forecasting might be due to the fact that the first autoregressive component is able to capture short-term fluctuations, but fails to do so for longer horizons.

The best performing quasi-AIM approaches in the first exercise are also among the best approaches when we analyze the different forecast horizons separately. For example, the quasi-AIM approaches with fixed effects and with or without the different September 11 coefficients across airports are among the best performers for short- and medium-term forecast horizons, while they perform relatively poorly for long horizons. The opposite is true for the quasi-AIM approach with just the different September 11 coefficients, which is among the worst for short- and medium-term forecasts and the best for long forecast horizons. These results suggest that allowing different constants across airports is relatively useful for short horizons, while exploiting the heterogeneity based on the September 11 reaction is useful for longer horizons. These results also provide us with a better understanding of the results of the first forecasting exercise, which is in some senses a combination of short-, medium- and long-term forecast horizons.

(a) Static forecasts.

(b) Short-term dynamic forecasts.

(c) Medium-term dynamic forecasts.
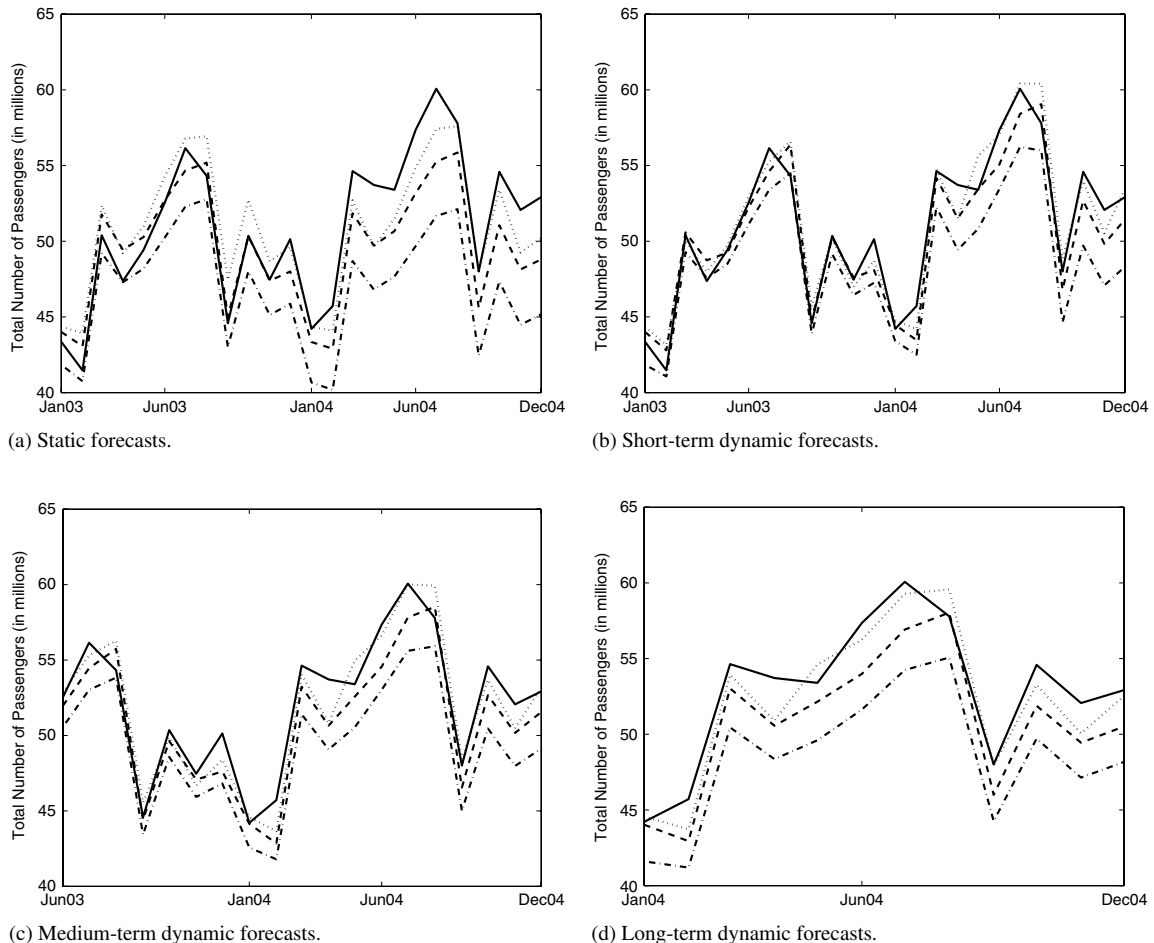

(d) Long-term dynamic forecasts.

Fig. 3. Aggregate US air travel demand and out-of-sample forecasts from selected approaches.
Note: The figure presents realized values for the aggregate US air travel demand, together with static forecasts (Panel (a)) and dynamic forecasts for short (Panel (b)), medium (Panel (c)) and long (Panel (d)) forecast horizons from selected forecasting approaches. In each panel, the solid line (——) represents the realized values for the aggregate US air travel demand; the dashed line (- - -) corresponds to out-of-sample forecasts from the AIM approach; the dotted line (· · ·) corresponds to out-of-sample forecasts from the quasi-AIM approach with fixed effects and different September 11 coefficients across airports; and the dash-dotted line (- · -) corresponds to out-of-sample forecasts from the aggregate approach. The out-of-sample period for the static forecasts is between January 2003 and December 2004, and the parameter estimates are based on the fixed window of observations between January 1990 and December 2002. The out-of-sample period for dynamic forecasts changes depending on the forecast horizon, and the forecasts are based on parameters estimated for an expanding window of observations. The out-of-sample periods for the short-, medium- and long-term forecast horizons are January 2003–December 2004, June 2003–December 2004 and January 2004–December 2004, respectively.

More importantly, independent of the forecast horizon, the aggregate approach is always the worst-performing approach, suggesting that there are always gains in terms of forecasting ability from using information which is available at the disaggregate level. Furthermore, almost all quasi-AIM approaches outperform the AIM approach for every forecast horizon, suggesting that quasi-AIM approaches might be better suited for exploiting the heterogeneity across airports, without suffering from as many of the problems associated with estimation uncertainty as the AIM approach. On the other hand, the same cannot be said about the pooled approach, as it is among the best approaches for long horizons, which might also explain its relatively good performance in the first forecasting exercise.

Fig. 3 presents realized values for the aggregate US air travel demand, as well as static and dynamic forecasts for the short-, medium- and long-term horizons from selected approaches. It allows us to verify the above findings about the out-of-sample performances of different approaches visually. For example, one can see from Fig. 3 that the quasi-AIM approach with fixed effects and different September 11 coefficients across airports generally outperforms the other approaches graphed in the figure.

Several other interesting facts emerge from Fig. 3. In the static forecasting exercise, all forecasting approaches tend to first overforecast then underforecast the aggregate air travel demand, with the exception of the aggregate approach. There are two possible explanations for this pattern of forecasting errors. First and more importantly, our forecasts of the explanatory variables are somewhat conservative. In other words, we underpredicted the explanatory variables ex-post, with significant positive effects on the per capita air travel demand, and in particular the variables related to economic growth. This is consistent with other available forecasts of economic activity during the time period considered in this paper. Another plausible reason might be related to the effect of the September 11 terrorist attacks. The aggregate air travel demand might have reacted more strongly to the September 11 terrorist attacks and recovered faster than our models predicted. Furthermore, as one would expect for most static forecasting exercises, the forecasting errors get larger with the forecasting horizon. In the second forecasting exercise, we attempt to address this problem by analyzing three different forecast horizons separately. Fig. 3 confirms our earlier claim that long-term forecasts generally underpredict the aggregate air travel demand, and the forecasting errors for short- and medium-term forecasts are relatively smaller than those for long-term forecasts.

To summarize, our out-of-sample results confirm our previous assertion that there are gains from using disaggregate level data when forecasting an aggregate variable. Moreover, employing the heterogeneity across airports yields superior forecasts. In other words, the disaggregate approaches (quasi-AIMs and AIM) outperform the aggregate approach. The question of which quasi-AIM approach performs the best depends on the source of most of the heterogeneity across markets, as well as on the forecast horizon.

6. Conclusion

In this paper, we analyze whether it is better to forecast air travel demand using aggregate data or to sum the airport-specific forecasts obtained from disaggregate data. We find that disaggregate forecasting approaches outperform the aggregate approach in terms of the mean absolute forecast errors and root mean square forecast errors of out-of-sample forecasts of the aggregate variable at different forecast horizons. We argue that the performance of a disaggregate approach depends on the trade-off between the degree of heterogeneity allowed and the estimation uncertainty. The AIM (aggregating individual markets) approach, which exploits all of the heterogeneity across individual airports, also suffers from high estimation uncertainty due to the number of coefficients estimated. We find that the approaches where we restrict the heterogeneity across airports by forcing the coefficient estimates to be the same across airports, which we term quasi-AIM approaches, outperform the AIM approach. We argue that the disaggregate approaches can be implemented in a straightforward manner, with the potential for better forecasts than either the aggregate or the AIM approach, since they exploit the heterogeneity in individual market dynamics efficiently, without high estimation uncertainty. More structural AIM variants can help to identify important individual market differences and provide inputs and links to choice models.

References

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.

Aigner, D. J., & Goldfeld, S. M. (1974). Estimation and prediction from aggregate data when aggregates are measured more accurately than their components. *Econometrica*, *42*, 113–134.

Allen, P. G., & Fildes, R. (2001). Econometric methods. In J. S. Armstrong (Ed.), *Principles of forecasting: a handbook for researchers and practitioners* (pp. 301–363). New York, Boston, Dordrecht, London, Moscow: Kluwer Academic Publishers.

Benalal, N., del Hoyo, J. L. D., Landau, B., Roma, M., & Skudelny, F. (2004). *To aggregate or not to aggregate: euro area inflation forecasting.* European Central Bank. Working Paper, Series No. 374.

Bronnenberg, B. J., Dhar, S. K., & Dube, J. (2007). Consumer packaged goods in the United States: national brands, local branding. *Journal of Marketing Research*, *44*, 4–13.

Cline, R. C., Ruhl, T. A., Gosling, G. D., & Gillen, D. W. (1998). Air transportation demand forecasts in emerging market economies: a case study of the Kyrgyz republic in the former Soviet Union. *Journal of Airport Management*, *4*, 11–23.

Energy Information Agency (2003). *Annual energy outlook*.

Espasa, A., Senra, E., & Albacete, R. (2002). Forecasting inflation in the European monetary union: a disaggregated approach by countries and by sectors. *European Journal of Finance*, *8*, 402–421.

Fair, R. C., & Shiller, R. J. (1990). Comparing information in forecasts from econometric models. *American Economic Review*, *80*, 375–389.

Giacomini, R., & Granger, C. W. J. (2004). Aggregation of space-time processes. *Journal of Econometrics*, *118*, 7–26.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics*, *14*, 227–238.

Granger, C. W. J. (1987). Implications of aggregation with common factors. *Econometric Theory*, *3*, 208–222.

Granger, C. W. J. (1990). Aggregation of time-series variables: a survey. In T. Barker, & M. H. Pesaran (Eds.), *Disaggregation in econometric modelling* (pp. 17–34). London, New York: Routledge.

Grunfeld, Y., & Griliches, Z. (1960). Is aggregation necessarily bad? *The Review of Economics and Statistics*, *42*(1), 1–13.

Hendry, D. F., & Hubrich, K. (2006). *Combining disaggregate forecasts or combining disaggregate information to forecast an aggregate.* European Central Bank. Working Paper, Series No. 589.

Ito, H., & Lee, D. (2005). Assessing the impact of the September 11 terrorist attacks on US airline demand. *Journal of Economics and Business*, *57*, 75–95.

Jordà, O., & Marcellino, M. (2010). Path forecast evaluation. *Journal of Applied Econometrics*, *25*(4), 635–662.

Kohn, R. (1982). When is an aggregate of a time series efficiently forecast by its past? *Journal of Econometrics*, *18*, 337–350.

Lehmann, D. R., & Winer, R. S. (2001). *Analysis for marketing planning*. McGraw-Hill.

Lutkepohl, H. (1984). Forecasting contemporaneously aggregated vector ARMA processes. *Journal of Business and Economic Statistics*, *2*(3), 201–214.

Lutkepohl, H. (1987). *Forecasting aggregated vector ARMA processes*. Springer-Verlag.

Lutkepohl, H. (2006). Forecasting with VARMA processes. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.), *Handbook of economic forecasting*. Elsevier.

Marcellino, M., Stock, J. H., & Watson, M. W. (2003). Macroeconomic forecasting in the euro area: country specific versus area-wide information. *European Economic Review*, *47*(1), 1–18.

Ord, K., & Young, P. (2004). Estimating the impact of recent interventions on transportation indicators. *Journal of Transportation and Statistics*, *7*(1), 69–87.

Pesaran, M. H., Pierse, R. G., & Kumar, M. S. (1989). Econometric analysis of aggregation in the context of linear prediction models. *Econometrica*, *57*, 861–868.

Profillidis, V. A. (2000). Econometric and fuzzy models for the forecast of demand in the airport of Rhodes. *Journal of Air Transport Management*, *6*, 95–100.

Saab, S. S., & Zouein, P. P. (2001). Forecasting passenger load for a fixed planning horizon. *Journal of Air Transport Management*, *7*, 361–372.

Theil, H. (1954). *Linear aggregation of economic relations*. Amsterdam: North Holland.

van Garderen, K. J., Lee, K., & Pesaran, M. H. (2000). Cross-sectional aggregation of non-linear models. *Journal of Econometrics*, *95*, 285–331.

Wang, P. T., & Pitfield, D. E. (1999). The derivation and analysis of the passenger peak hour: an empirical application to Brazil. *Journal of Air Transport Management*, *5*, 135–141.

Zellner, A., & Tobias, J. (2000). A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*, *19*, 457–469.