Dynamic modelling of fares and passenger numbers for major U.S. carriers

Anthony Martin, Maximilian Martin, Steve Lawford

Project: Dynamic modelling of fares and passenger numbers for major U.S. carriers *

ANTHONY MARTIN and MAXIMILIAN MARTIN
(supervised by STEVE LAWFORD)†

Department of Economics and Econometrics, ENAC, France.

March 13, 2010

Abstract

The purpose of this project was to develop econometric models that will enable us to describe and forecast the evolution of air fares and passenger numbers for the 7 largest U.S. carriers. The principal data source was the Department of Transport's DB1B database, which contains extensive information on airline tickets sold in the US. The modelling was first conducted on the basis of pure statistical models, and we later introduced variables with real economic data, such as the air carrier's financial situation and data on the U.S. economy and also information on the individual route characteristics. Among other results, the statistical models reveal a significant drop in fares, and often also passenger numbers, after the September 11 attacks as well as a visible difference between the behaviour of Southwest Airlines , the only low-cost airline under examination, and the other six major carriers. The selection and assessment of the models in this project has been fully automated, using original code developed by the authors in the ecenometric software package EViews6 and with Excel-VBA.

---

*Journal of Economic Literature (JEL) classification: C01 Econometrics, C13 Estimation, C22 Time-series models, C23 Models with panel data, C51 Model construction and estimation, C53 Forecasting and other model applications, C80 Data collection and data estimation methodology; computer programs - general, L10 Market structure, firm strategy, and market performance, L93 Air transportation (see http://www.aeaweb.org/journal/jel class system.html for classification system). Keywords: airline strategy, ARIMA models, forecasting tools, time series and panel data, U..S. major carriers.

†Steve Lawford, LH/ECO, ENAC, 7 avenue Edouard Belin, BP 54005, 31055, Toulouse, Cedex 4, France. Email: steve_lawford (at) yahoo.co.uk.

1 Introduction

The aim of this project is to develop econometric models that are capable of describing and forecasting the evolution of air fares and passenger numbers in the U.S. domestic market, for the 7 largest U.S. carriers, American Airlines, Continental Airlines, Delta Airlines, Northwest Airlines, Southwest Airlines, United Airlines and US Airways. The main data source used here is the official DB1B database which holds information on a large representative sample of tickets sold in the U.S. from 1993 to 2009. For certain models we also use data collected from the airlines by the Bureau of Transportation Statistics (the Form 41 Data), and data on the economic situation of the U.S. economy from the Federal Reserve Bank of St. Louis.

The first step is to find the best model that is based purely on statistical data. We examine several auto-regressive (integrated) moving average (AR(I)MA) models for both fares and passenger numbers and identify the best model, which we validate and then use for short-term forecasting. In a second step, we take the best AR(I)MA models and try to improve them by using data on the U.S. economy, that is the GDP and data on the economic situation of the airlines such as fuel expenses and operating profits. We then apply the developed models to illustrate routes, and also examine the airline's behaviour using a panel dataset and additional economic and market/route-specific characteristics.

The structure of this project is as follows. Section 2 provides background information and key features of the seven carriers of interest. Section 3 presents the core Department of Transport datasets, the filtering assumptions that were made, and the method used to deflate nominal in real prices. In section 4, we introduce the Box-Jenkins methodology for building statistical ARIMA time-series models. This is applied to each of the carriers in turn, the models are carefully validated, and used for short-term and dynamic forecasting. Section 5 discusses additional economic variables that are used in an effort to improve the ARIMA models. Section 6 applies the modeling to several illustrative routes. In Section 7, we build a number of purely economic models for fares, which are used for comparison and to aid intuition. Section 8 concludes the project. The Appendix contains automated model selection output, the EViews 6 and Excel VBA code that was *entirely written by the authors* during the project (for data treatment and analysis), and several Python codes for initial treatment (cleaning, formatting, etc.) of the DB1B and economic data. The report was written and compiled using the typesetting software WinEdt and the LaTeX language.

Contents

2 Major U.S. airlines

2.1 American Airlines

American Airlines (IATA code AA) is one of the world's largest airlines (in terms of traffic and operating revenues).[1] Established in 1930 in New York, the airline initially operated scheduled services between Boston, New-York, Chicago, Dallas and Los Angeles. In 1979, one year after the U.S. deregulation of the domestic network, American Airlines moved its headquarters to Fort Worth in Texas.[2] As a major carrier (defined as having more than $1 billion in annual operating revenues), the airline chose to operate a hub-and-spoke network, with 3 main bases in San Jose (California), Raleigh (North Carolina) and Nashville (Tennessee). During the 1980s and 1990s, American Airlines largely expanded its network, throughout the domestic U.S., and internationally. In 1999, the airline created the OneWorld global airline alliance with British Airways, Canadian, Cathay Pacific and Quantas.[3] In April 2001, American purchased the assets of the near-bankrupt Trans World Airlines (TWA), at the same time acquiring its hub in Saint Louis. The September 11, 2001, terrorist attacks (2 of which involved American Airlines aircraft) had serious consequences on the financial situation of the company.[4] In July 2005, American Airlines announced its first quarterly profit since September 2001. American Airlines is the only legacy carrier to survive bankruptcy-free.

American Airlines is a subsidiary of the AMR corporation. Since 2004, the Chairman, President, and Chief Executive Officer of AA and AMR has been Gerard Arpey. The airline is now based in

[1] See the 2009 economic report of the Air Transport Association. Refer to [2].

[2] For further information on U.S. deregulation of domestic traffic, refer to [4].

[3] This alliance was started by 5 of the world's largest airlines at the time, with a driving strategy of expansion of the international network of each of these airlines, through code-sharing and then an immunized alliance. For American Airlines in particular, it was an opportunity for more efficient cooperation on Transatlantic routes with British Airways and on Pacific routes with Quantas and Cathay Pacific.

[4] In fact, American airlines managed to reduce the immediate impact of the September 11 attacks. As a consequence the airline did not file for bankruptcy, whereas Delta, Northwest, United and US Airways did and consequently benefited from the help of the federal government. Therefore the airline found itself at a competitive disadvantage compared to these 4 major airlines. See also [3].

Dallas/Fort Worth and has 2 main bases at Chicago O'Hare International and Miami International.

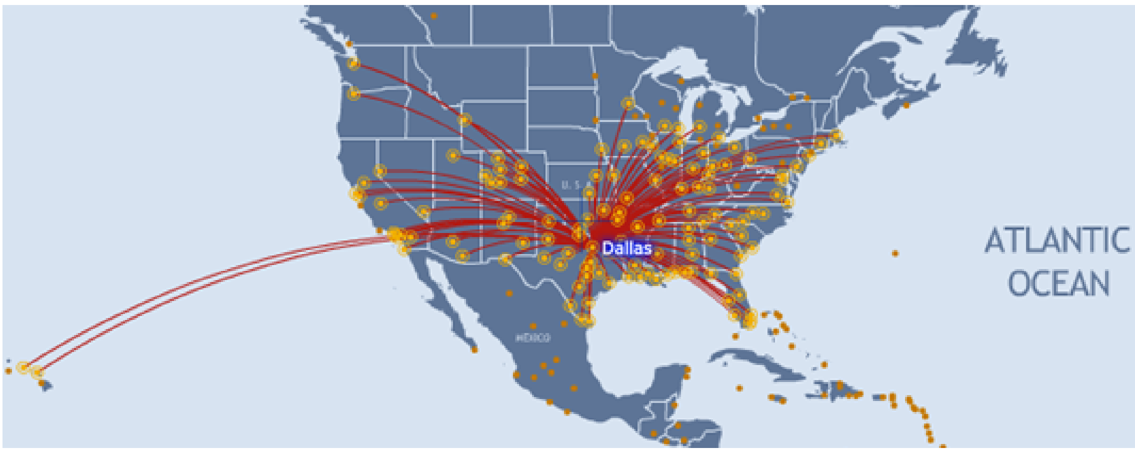


Figure 1: Domestic traffic of American Airlines from Fort Worth International Airport (ENAC Air Transport Data).

2.2 Continental Airlines

Continental Airlines (IATA code CO) is the fourth largest U.S. airline in terms of operating revenues and revenue passenger miles (RPMs). Established in 1934, and based in Houston (Texas), Continental offers domestic, regional and international flights. Created by Walter Varney (who also founded United Airlines), Continental Airlines was originally a regional airline, smaller (in terms of number of routes) than TWA, American Airlines and United. Under its CEO Robert Six, the airline expanded significantly between 1959 and 1978, and was the first carrier to introduce the Boeing 707 and then the Boeing 747 on its domestic routes. Six was also the first to introduce an economy fare on the Chicago-Los Angeles route. Continental offered services from Denver to Chicago, Seattle, Los Angeles and Honolulu. After deregulation, the carrier continued to expand significantly, notably towards the Eastern coast of the U.S., into Miami and New York. However, its acquisition by the airline holding company Texas Air, and a series of reorganizations, led to a first Chapter 11 bankruptcy filing from 1983 to 1986.[5] In 1988, Continental Airlines formed the first international alliance with the Scandinavian airline SAS. From 1990 to 1993, the airline entered its second bankruptcy. Hired in 1993 in order to salvage the company, the new CEO of Continental, Gordon Bethune, began to focus more on customer satisfaction than on fares, and made the airline in 2004 the "most admired global airline" according to Fortune magazine.[6] In the same year, Continental entered the SkyTeam alliance between Air France and Delta Airlines, along with Northwest and KLM. This alliance was initiated by Northwest, which at that moment had the control over all decisions taken by Continental. After the acquisition of Northwest by Delta Air Lines in 2008, Continental Airlines decided to withdraw from the SkyTeam alliance and join the Star Alliance in order to cooperate with United Airlines.[7] This change of alliance was effectively completed in October 2009.

The CEO of Continental Airlines since 2004 has been Larry Kellner, following the retirement of Gordon Bethune. Since the latter was a Boeing executive before managing the airline, from 1994 Continental Airlines has began to renew its fleet only with Boeing aircraft. The three main bases of Continental Airlines are Houston, New York (Newark) and Cleveland.

2.3 Delta Air Lines

Delta Air Lines (IATA code DL), established in 1934, is currently the main U.S. major carrier, in terms of traffic and operating revenues. The airline was created in Macon, Georgia, in 1924, as part of the company Huff Daland Dusters. The headquarters was moved to Atlanta in 1941. Beginning

[5] Chapter 11 is a section of the United States Bankruptcy Code, which allows a company to reorganize rather than be liquidated. This reorganization is performed by the debtor, to whom the U.S. government provides mechanisms for it to restructure its business, especially the ability to acquire loans and financing from new lenders without clearing its previous debts.

[6] Refer to [1].

[7] Indeed, Continental had several merger discussions with United Airlines between 2006 and 2008, and the CEO of Continental stated on November 2009 that "We are watching Delta to see whether Delta outperforms us financially" before reconsidering the merger with United [7].

Figure 2: Domestic traffic of Continental Airlines from George Bush Intercontinental Airport (ENAC Air Transport Data).

its passenger services on the West coast, the acquisition of Northeast Airlines in 1972 greatly helped the expansion of the carrier in the Eastern U.S. In 1955, it was the first airline to implement the hub-and-spoke system. In 1987, Delta merged with Western Airlines, acquiring the hubs of Salt Lake City and Los Angeles (Los Angeles International Airport). However, the main acquisition of Delta Airlines was its absorption of the bankrupt airline PanAm in 1991. In 2000, the carrier founded the SkyTeam alliance with Air France, Aeromexico and Korean Air. Four years later, Delta began an important restructuration of its services in an effort to improve profitability and avoid bankruptcy. Despite these measures, the airline filed for Chapter 11 bankruptcy protection in September 2005, and returned as an independent carrier in April 2007. In October 2008, Delta Air Lines and Northwest Airlines (which was part of the SkyTeam alliance since 2004) merged under the Delta identity.

The CEO of Delta is Richard Anderson. As for Continental Airlines, Delta only operates with

Boeing aircraft.[8] The carrier, based in Atlanta, has 3 main U.S. bases: Cincinnati, New York (JFK Airport) and Salt Lake City.

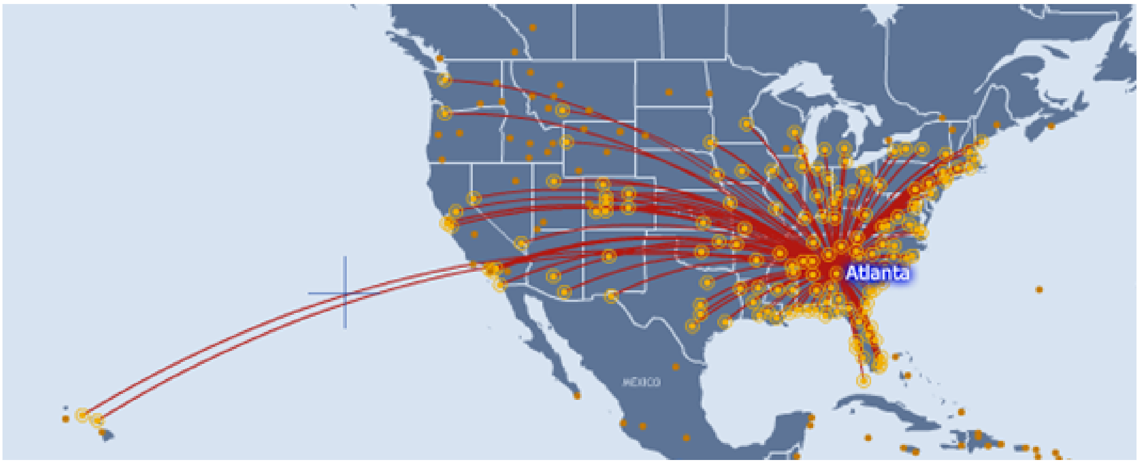


Figure 3: Domestic traffic of Delta Air Lines from Hartsfield-Jackson Atlanta International Airport (ENAC Air Transport Data).

2.4 Northwest Airlines

At the time of writing, Northwest Airlines (IATA code NW, established 1926) no longer exists and the Northwest brand name will be phased out by 2010. However, as its merger with Delta Airlines only took place in 2008, for the largest part of the time-frame in this study Northwest Airlines existed as an independent carrier. As for most U.S. majors, Northwest has more than 50 years of history, that began at its present headquarter city Minneapolis/St.Paul. Besides Minneapolis/St.Paul, Northwest operates two other U.S. hubs in Detroit and Memphis. Outside of the U.S., Northwest has a strong presence at Tokyo Narita and Amsterdam. Amsterdam is the hub airport for KLM Royal Dutch Airlines, with which Northwest pioneered the field of airline alliances/cooperations by entering a joint-venture that includes revenue sharing and Frequent Flyer

[8] In 1997, Delta announced that the airline would buy only Boeing aircraft for the next 20 years, in order to strengthen cooperation and as a consequence the flexibility and price advantages following from this [5].

Program cooperation. Today, Northwest is, as its "predator" Delta, member of the SkyTeam alliance. Northwest's fleet ranges from DC9-30s with 100 seats to 747-400s with 403 seats. All the aircraft are operated in a 2-class configuration with Economy and Business or First class. In 2008 it transported 65.5 million passengers and employed 34,400 people. Its net loss lay at $6.2 billion.

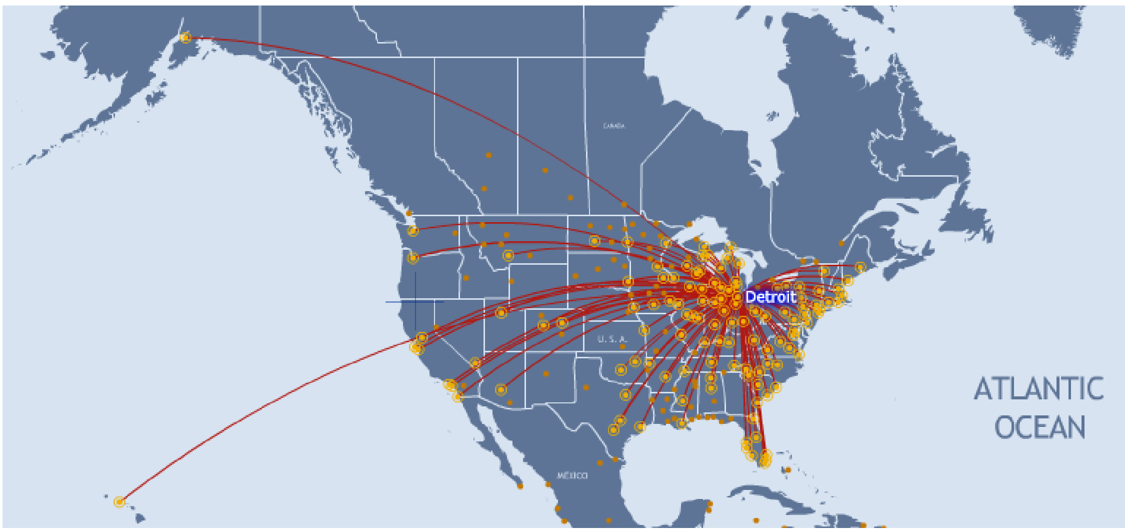


Figure 4: Domestic traffic of Northwest Airlines from Detroit Metro Airport (ENAC Air Transport Data).

2.5 United Airlines

United Airlines (IATA code UA, established 1931) has developed to become one of the largest airlines in the U.S. and also worldwide. In 1997 United was, together with Air Canada, Lufthansa, SAS Scandinavian Airlines and Thai Airways, a founding member of Star Alliance, which is now the world's largest airline alliance. According to the Air Transport Association of America's 2008 Annual Report, United Airlines was third largest carrier by revenue passenger miles (RPM). At 109.8 billion RPM it was about 20 billion RPM behind American Airlines and 30 billion ahead of Continental Airlines. The airline is headquartered in Chicago, and operates hubs at Chicago

O'Hare, Washington Dulles, San Francisco and Denver (the latter is purely domestic). From its hubs it serves destinations on all continents except Africa. The first African destinations will be Accra and Lagos in 2010. United operates 360 aircraft (2008) ranging from the Airbus A319 to the Boeing 747-400 in a 3- or 4-class-layout. Its Economy product has recently been supplemented at the top end and below Business/First Class with the introduction of the so-called "Economy Plus" class which is available on all United aircraft. Several smaller airlines fly under the United Express brand name to supplement United's domestic network.

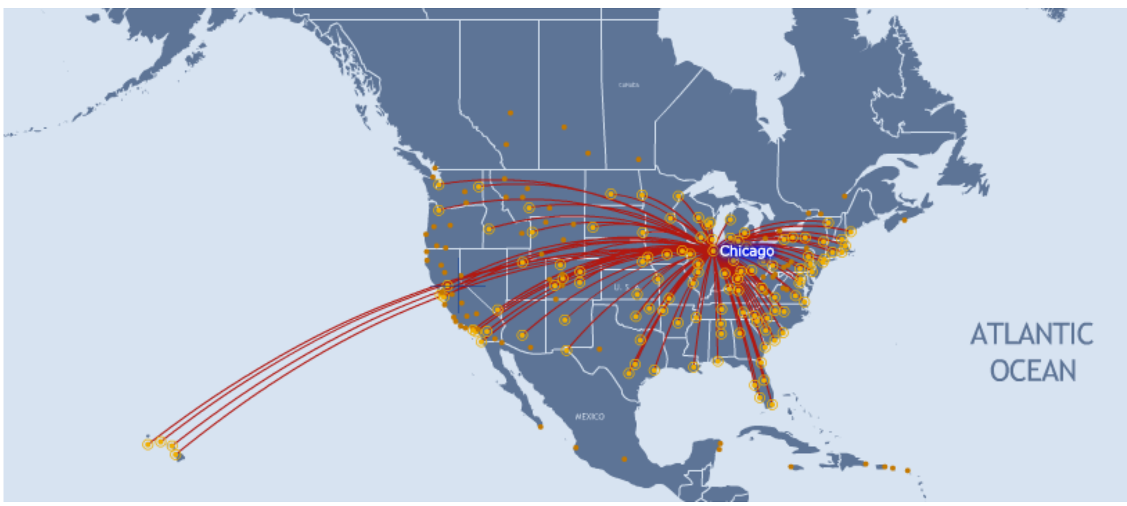


Figure 5: Domestic traffic of United Airlines from Chicago O'Hare International Airport (ENAC Air Transport Data).

2.6 US Airways

US Airways (IATA code US, established 1939) is the smallest of the airlines we will look at in this project. In 2008, its transport performance lay at 60.5 million RPM, trailing Southwest, and it employed 32,700 people while posting a loss of $2.2 billion. Before its merger with America West in 2005, US Airways mainly served the eastern part of the U.S. from its hubs in Philadelphia and Charlotte. The 2005 merger with America West added America West's Phoenix hub to the network.

Since 2004, US Airways has been a member of Star Alliance and operates 349 aircraft (296 aircraft fly with other airlines such as Mesa Air and PSA Airlines under the US Airways Express brand). All aircraft are equipped with an Economy and Business or First class, the latter called “Envoy” class. The airline is headquartered in Phoenix and mainly serves destinations in North America and Europe.

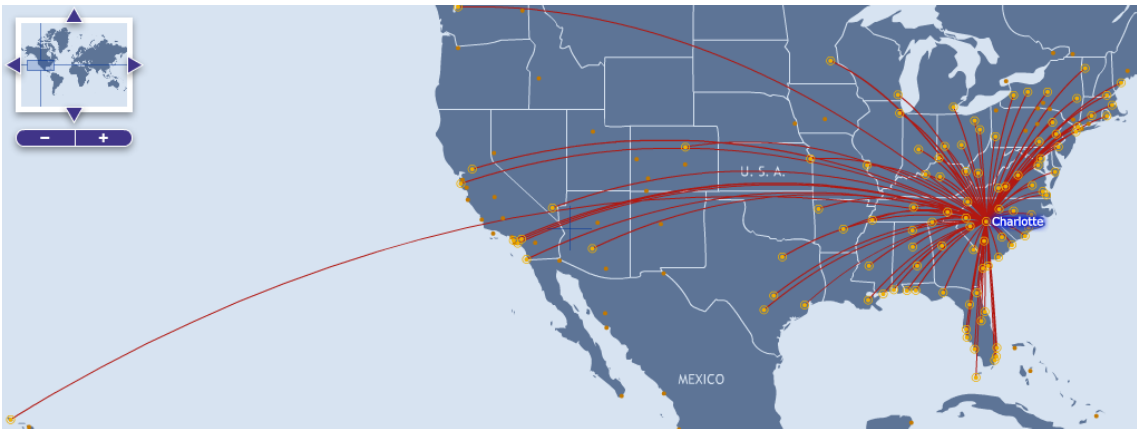


Figure 6: Domestic traffic of US Airways from Charlotte Douglas International Airport (ENAC Air Transport Data).

2.7 Southwest Airlines

Southwest (IATA code WN, established 1967) is the only low-cost carrier considered in this project. However, even though it is low-cost and also the youngest among the carriers we examine, Southwest is the largest by number of passengers transported (88.5 million in 2008). It is quite different compared to the other large U.S. carriers. For example, Southwest Airlines posted a net profit ($0.18 billion) in 2008. The other main differences apart from its age and low-cost/low fare concept are that it has a single-class 737-only fleet (namely 181 737-300 and 338 737-700 with 137 seats each and 25 737-500 with 122 seats) and operates exclusively on U.S. domestic routes. Southwest is not part of any larger alliance, and only has agreements with Canadian carrier Westjet and Mexico’s

Volaris. Southwest is headquartered in Dallas. Its main bases are Las Vegas, Chicago, Phoenix, Baltimore/Washington, Houston, Dallas, Denver, Oakland, Los Angeles and Orlando.

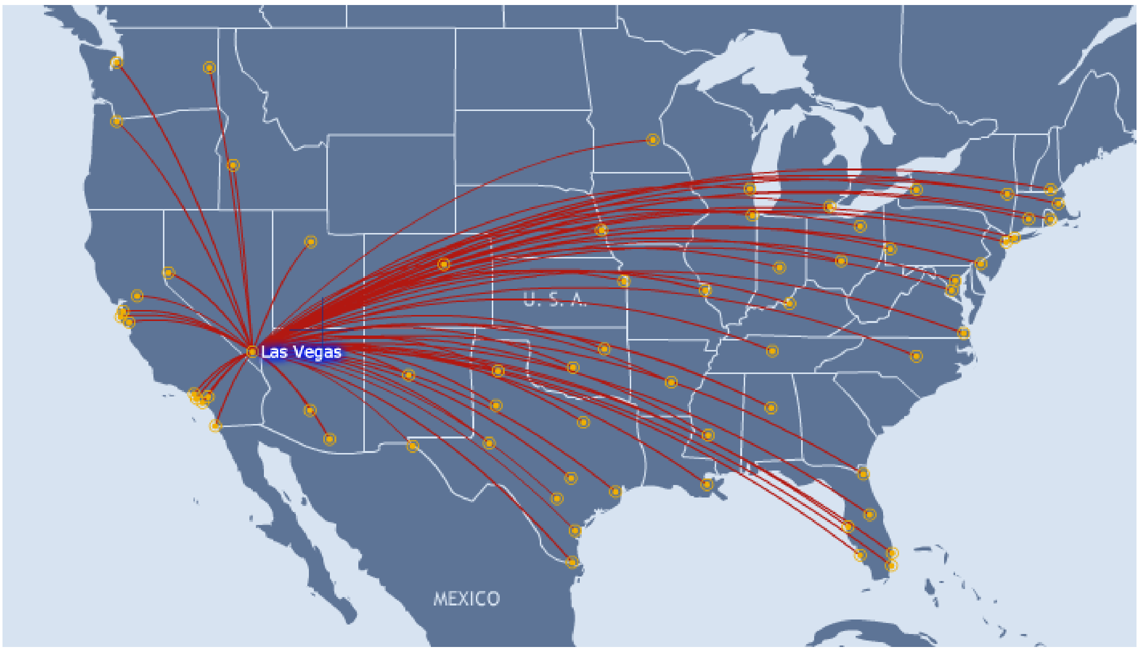


Figure 7: Domestic traffic of Southwest Airlines from Las Vegas McCarran International Airport (ENAC Air Transport Data).

3 Data description and preliminary analysis

3.1 Core data

In this project, we will make use of a number of rich statistical data sets from the American Bureau of Transportation Statistics (BTS).[9] We are particularly interested in the BTS Airline Origin and Destination Survey (DB1B). This dataset is a 10% random sample of scheduled airline tickets from reporting carriers. It gives information on domestic itineraries within the U.S. domestic market. The data has been reported quarterly from 1993 to 2009 (we have access to reliable data until the first quarter of 2009), giving us data over 65 quarterly time periods.

The survey is divided into 3 related databases: Market, Coupon, and Ticket. We will only use the second and third for this project. In fact, the Market database mostly uses the data contained in the Coupon and Ticket databases to report a new presentation of itineraries, displaying the number of passengers, the fares, the time period, the ticketing, operating and reporting carriers and the origin and destination airport for each segment of the flight.

3.1.1 DB1B Ticket data

This database gives, for each domestic itinerary reported by a reporting carrier, the origin and destination of the flight, the nominal itinerary fare, the number of passengers per ticket, and the length of the trip, among other variables. Figure 8 presents a small excerpt from this database, highlighting the main variables that are of use to us.

- **ItineraryID (ItinID):** This variable links the DB1B databases Coupon and Ticket, and allows us to compare and merge the data. One ticket has a unique "ItinID", and it can contain several coupons (segments) with that same ID.

[9] This agency, created in 1992, has been collecting data from any mode of transportation for more than 15 years, creating huge microeconomic and other databases in many areas (infrastructure, economic, financial, social, demographic, energy, freight or passenger travel). Much data is freely available online.

ItineraryID	Number of Coupons	Year	Quarter	Origin Airport	Origin state	Reporting Carrier	Passengers	Itinerary Fare	Distance
200912901646	1	2009	1	FLL	FL	WN	2.00	215.00	793.00
200912901647	1	2009	1	FLL	FL	WN	1.00	218.00	793.00
200912901648	1	2009	1	FLL	FL	WN	2.00	221.00	793.00
200912901649	1	2009	1	FLL	FL	WN	6.00	227.00	793.00
200912901650	1	2009	1	FLL	FL	WN	3.00	248.00	793.00
200912901651	1	2009	1	FLL	FL	WN	1.00	250.00	793.00
200912901652	1	2009	1	FLL	FL	WN	2.00	258.00	793.00
200912901653	1	2009	1	FLL	FL	WN	3.00	260.00	793.00
200912901654	1	2009	1	FLL	FL	WN	1.00	272.00	793.00
200912901655	1	2009	1	FLL	FL	WN	1.00	31.00	793.00
200912901656	2	2009	1	FLL	FL	WN	1.00	117.00	1916.00
200912901657	2	2009	1	FLL	FL	WN	1.00	148.00	1645.00
200912901658	2	2009	1	FLL	FL	WN	2.00	272.00	970.00
200912901659	2	2009	1	FLL	FL	WN	1.00	287.00	970.00
200912901660	4	2009	1	FLL	FL	WN	1.00	267.00	1627.00
200912901661	4	2009	1	FLL	FL	WN	1.00	302.00	4980.00
200912901662	2	2009	1	FLL	FL	WN	1.00	5.00	1381.00
200912901663	2	2009	1	FLL	FL	WN	1.00	88.00	1381.00
200912901664	2	2009	1	FLL	FL	WN	2.00	151.00	1381.00
200912901665	2	2009	1	FLL	FL	WN	1.00	164.00	1381.00
200912901666	2	2009	1	FLL	FL	WN	1.00	173.00	1381.00
200912901667	2	2009	1	FLL	FL	WN	1.00	214.00	1381.00

Figure 8: Excerpt from the DB1B Ticket database for 2009Q1, where FLL is Fort Lauderdale Airport (IATA airport code), FL is Florida, WN is Southwest (IATA airline code), the itinerary Fare is nominal (in U.S.$), and distance is measured in miles. In the text, we take a closer look at the 3 highlighted itineraries.

- **Number of Coupons:** It is essential to know, for a given itinerary, the number of segments contained in the itinerary. The minimal number of coupons is 1, and reaches 23. Figure 9 shows, for example, the distribution of the number of coupons bought by passengers on an itinerary in 2008. We can see that the proportion of tickets with 7 coupons or more seems negligible (it represents, in 2008, 0.25% of all tickets sold).

- **Year and Quarter:** Tickets are reported every quarter, and it would be very difficult to obtain the exact date of the flight for each ticket (this is not available through DB1B), which would add considerable complexity to the databases, although it would enable us to model for example intraweek, weekly and monthly effects. From 1993Q1 to 2009Q2, there are consequently 65 different periods in the database.

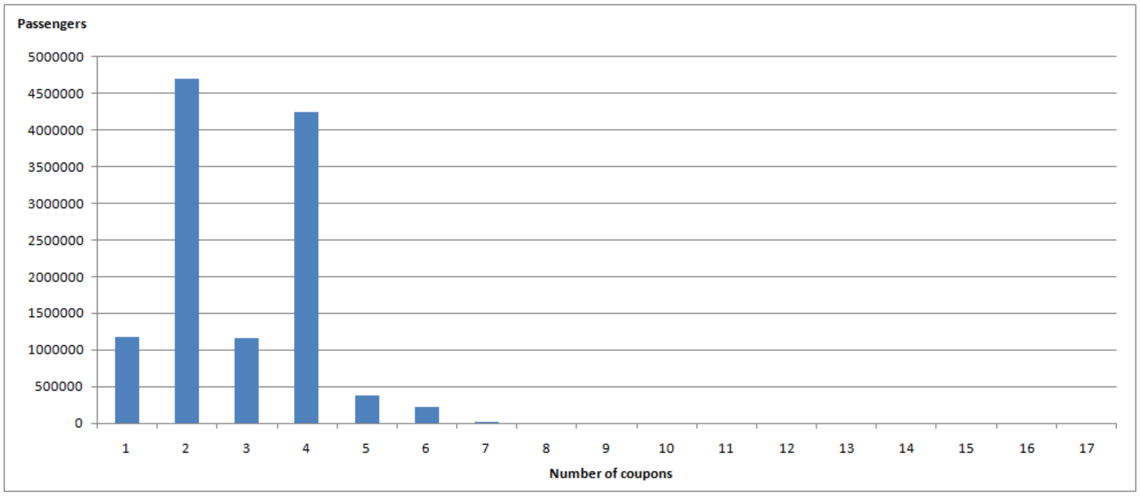


Figure 9: Coupons bought by passengers on an itinerary in 2008, for all U.S. reporting carriers. We observe that the proportion of tickets with 2 or 4 itineraries represents more than 66% of all tickets sold. In fact, this graph does not give us the proportion of round trips or changes of planes, so we cannot state whether the 4,250,000 4-coupon itineraries are one-stop returns or one-way trips with 4 successive changes of plane.

Consider the number of passengers who took a flight during the year 2008 (Figure 10). We can observe that the quarter when people travelled the most was the second quarter, with 500,000 more travellers than during the first quarter of 2008. About 2,800,000 passengers travelled during 2008Q1 and a similar number during 2008Q4.

- **Origin airport and state:** We have the origin airport for each ticket. The DB1B Coupon dataset will later provide the entire route followed by each passenger. Figure 11 shows, for example, the distribution of the passengers having a trip in 2008 from a given state of origin. We can see that the outgoing traffic comes mostly from California, Texas, New York, Florida, Illinois and Pennsylvania, which are also the most populous U.S. states.[10] As a consequence, we could try to avoid this effect of the most populous states on the traffic results, by dividing the number of passengers for a given state by the population of this state (Figure 12). We

[10] 2007 population estimates from the U.S. Census Bureau. See `http://www.census.gov/`.

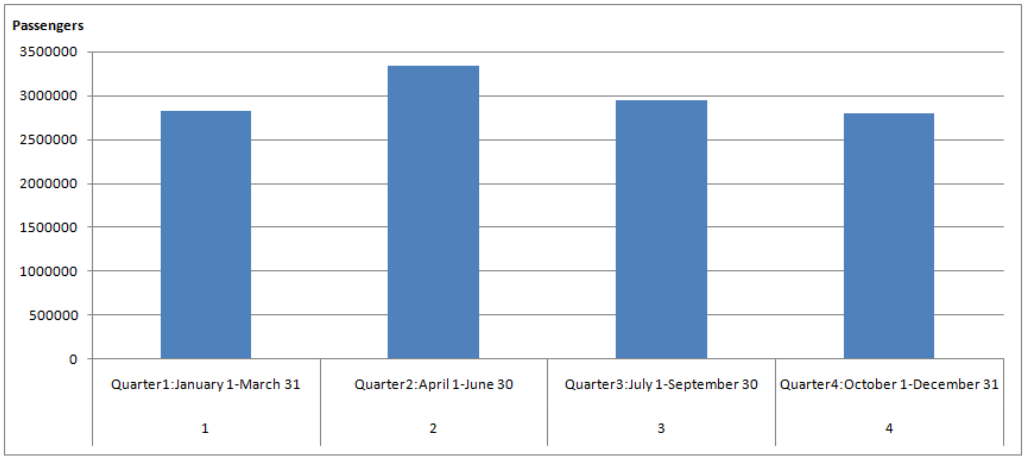


Figure 10: Number of passengers per quarter in 2008.

can see from this graph that this kind of concentration of the traffic is very high for the less populous U.S. states, especially for the District of Columbia. In 2008, a number of passengers equivalent to half the population of Washington D.C. left the capital of the United States for domestic travel. We can also note the high percentage for Alaska, which can be explained by the low population of the state and the difficulty for the inhabitants to make domestic trips in the U.S. with other means of transportation than aircraft. The same arguments and the high level of tourism for Hawaii can explain the high percentage calculated for the island. For Nevada, the low population compared to the high traffic from Las Vegas and Reno also explains the high percentage for the state. Returning to the first graph, we can also note the very low outgoing traffic in Dakota, Wyoming and West Virginia, which are also among the less populated U.S. states.

- **Reporting carrier:** The reporting carrier is the airline reporting the data of tickets and coupons to the Bureau of Transportation. We can examine the distribution of the passengers in 2008 by respective reporting carrier (Figure 13). We note that the largest reporting airlines (in terms of number of passengers) are the 6 legacy carriers presented in the first part of this

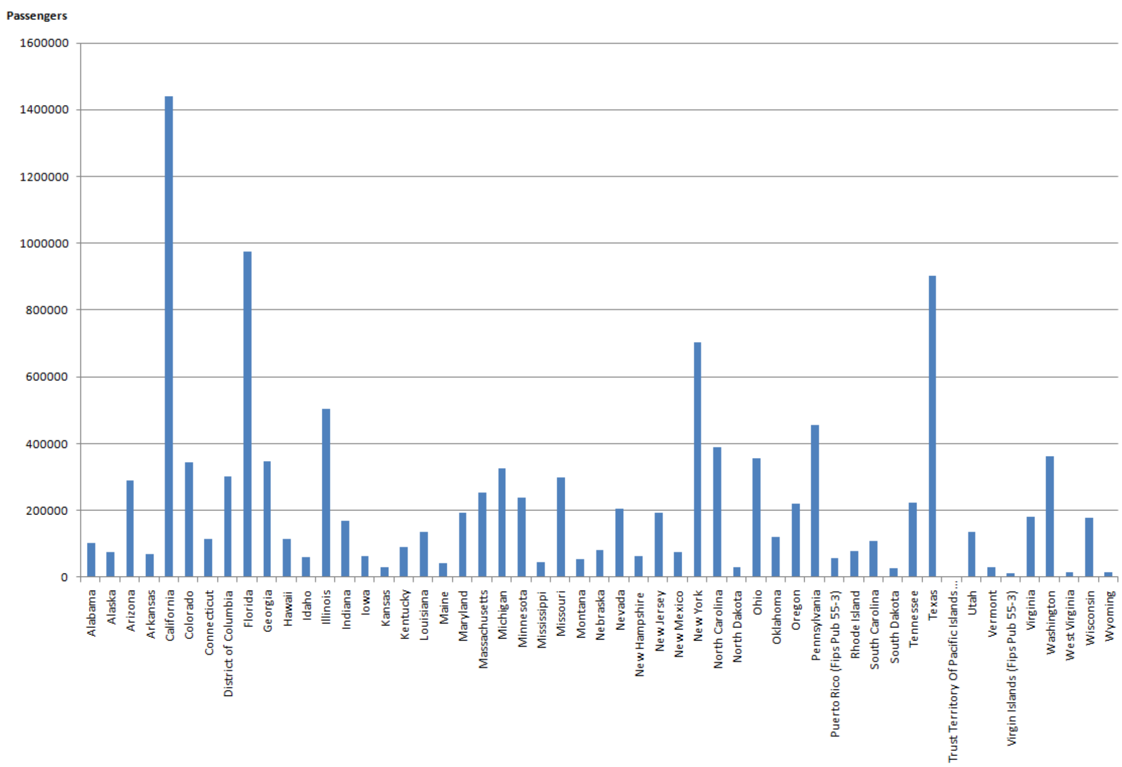


Figure 11: State of departure for 2008 itineraries.

project, and the low-cost carrier Southwest. These airlines are major carriers (as defined above), and reported the tickets of more than 600,000 passengers in 2008. The largest carrier on this graph is Southwest: we could consequently tend to think that the low-cost carrier was the most important (reporting) carrier for domestic travel in the U.S. in 2008. Nevertheless we have to keep in mind that the reporting carrier may not be the operating carrier, especially when a carrier like the 6 other legacy carriers uses code-sharing to develop its hub-and-spoke network. The code-sharing effects are very difficult to capture here, as the different alliances between airlines can change over time.

We also observe that there are other significant reporting carriers in the U.S.: SkyWest,

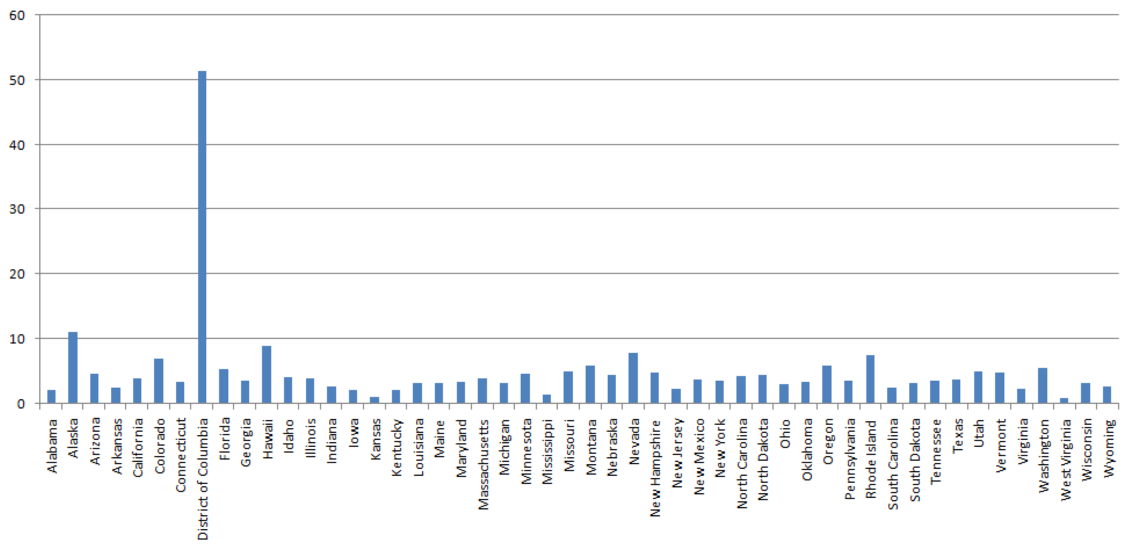


Figure 12: Number of passengers for 2008 itineraries as a percentage of the population estimates by U.S. state.

AirTran and American Eagle. SkyWest is a regional airline headquartered in Utah, and has recently secured a partnership with the low-cost carrier AirTran. This partnership will be effective in February 2010, and may strengthen the position of these 2 airlines in the U.S. domestic market. American Eagle is a subsidiary of American Airlines.

- **Passengers:** The number of passengers buying a ticket for a given route at a given price with a given reporting carrier. This number can be sometimes very high (up to 1000 passengers with the same ItinID: it does not mean that all these passengers know each other and bought one ticket for 1000 individuals, but instead that 1000 passengers bought a similar ticket during the quarter (the same route, with the same carrier(s), and at the same price), which enables the DOT to compress the dataset to some extent.

- **Itinerary fare:** This is the nominal U.S.$ fare paid for the ticket. We saw, looking at the database, that some fares appear unusual, as they seem to be very high for domestic travel, or

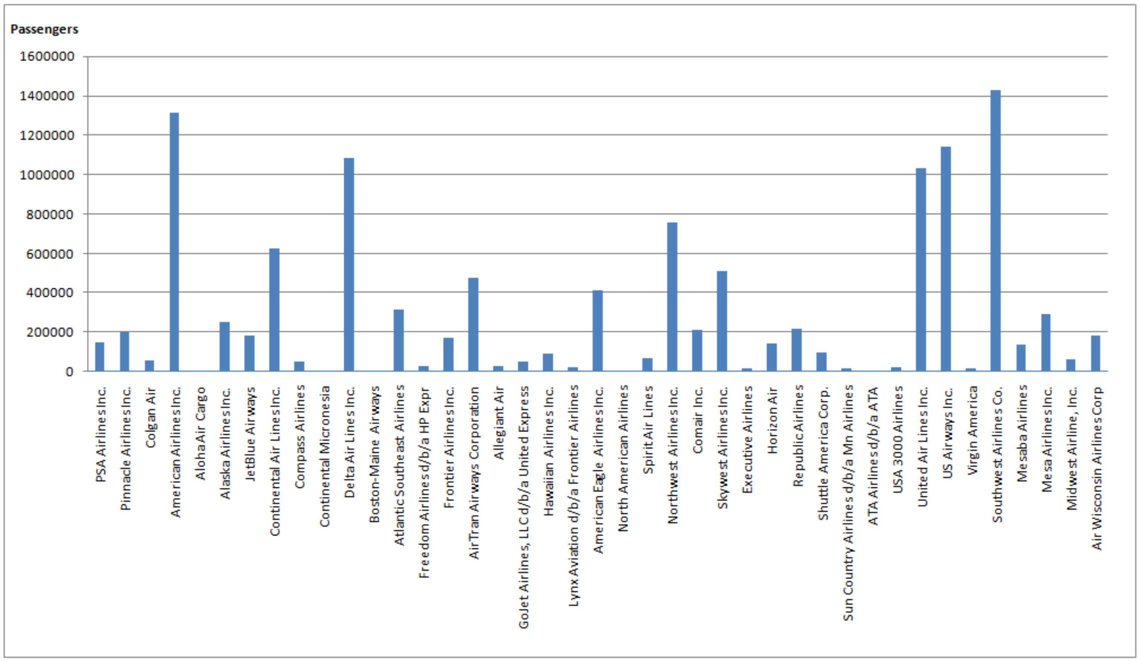


Figure 13: Carrier reporting 2008 itineraries.

very low (for example, 7 euros for a 3-coupons itinerary from Texas with Southwest). Indeed a passenger, even flying with a low-cost carrier such as Southwest, must pay some fees, charges and taxes: in order to remove possible errors from the database, we decided to impose upper and lower bounds on the fare data. We develop the computation of these bounds in section 3.1.3 below (Assumptions).

- **Distance:** The length of the route is expressed in miles. It may be interesting to see whether we can relate this distance to the fares proposed by the airlines on specific routes.

We now take a general look at 3 illustrative itinerary IDs (highlighted in red on Figure 8).

- **ItinID 200912901650:** This line represents one ticket, an itinerary of 1 (number of coupons) flight which took place during the first quarter of 2009. This itinerary began at Fort Lauderdale Hollywood International Airport (FLL), in Dania Beach, in the South of Florida (FL). This ticket was reported by Southwest to the Bureau of Transportation. On this itinerary, 3 passengers bought a ticket at a nominal fare of $248 per passenger, and we see that they covered 793 miles.

- **Itin 200912901658:** Starting from the same airport (FLL), in the same quarter, two other passengers performed 2 flights (perhaps a round trip), covering an itinerary of 970 miles which cost nominal $272 per passenger. This ticket was also reported by Southwest.

- **Itin 200912901660:** Another passenger, with the same initial conditions of time and place, covered 1627 miles with 4 successive flights (perhaps a non-direct round trip). According to the reporting carrier Southwest, this passenger paid a nominal fare $267 for this ticket.

3.1.2 DB1B Coupon data

We now consider the Coupon database. A small excerpt from the database (Figure 14) follows. As for the Ticket database, we are particularly interested in the following variables:

- **ItineraryID:** See Ticket data description above.

- **Sequence number and Coupons:** For a given ticket, there can be several coupons, which specify the number of separate flights in an itinerary.

- **Year and quarter:** See Ticket data description above.

- **Origin/Destination airport and state:** In the Coupon database, we again find the initial airport of departure for the first coupon of an ItinID, which must be the same as the Origin airport for the respective ticket. We can also see all the airports where the passenger landed, and can therefore reconstitute the whole route.

- **Break:** This variable indicates whether the passenger stopped at a given airport (if the variable is checked “x”) for a few hours or days. Thus if there is no break at a given destination

Itinerary ID	Sequence number	Coupons	Year	Quarter	Origin Airport	Origin State	Destination Airport	Destination State	Break	Ticketing carrier	Operating carrier	Reporting carrier	Passengers	Fare Class	Distance
200912901648	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	2.00	G	793.00
200912901649	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	6.00	G	793.00
200912901650	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	3.00	G	793.00
200912901651	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	1.00	G	793.00
200912901652	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	2.00	G	793.00
200912901653	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	3.00	G	793.00
200912901654	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	1.00	G	793.00
200912901655	1	1	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	1.00	X	793.00
200912901656	1	2	2009	1	FLL	FL	BNA	TN		WN	WN	WN	1.00	G	793.00
200912901656	2	2	2009	1	BNA	TN	ABQ	NM	X	WN	WN	WN	1.00	G	1123.00
200912901657	1	2	2009	1	FLL	FL	BNA	TN		WN	WN	WN	1.00	G	793.00
200912901657	2	2	2009	1	BNA	TN	BDL	CT	X	WN	WN	WN	1.00	G	852.00
200912901658	1	2	2009	1	FLL	FL	BNA	TN		WN	WN	WN	2.00	F	793.00
200912901658	2	2	2009	1	BNA	TN	BHM	AL	X	WN	WN	WN	2.00	F	177.00
200912901659	1	2	2009	1	FLL	FL	BNA	TN		WN	WN	WN	1.00	F	793.00
200912901659	2	2	2009	1	BNA	TN	BHM	AL	X	WN	WN	WN	1.00	F	177.00
200912901660	1	4	2009	1	FLL	FL	BNA	TN	X	WN	WN	WN	1.00	G	793.00
200912901660	2	4	2009	1	BNA	TN	BHM	AL		WN	WN	WN	1.00	G	177.00
200912901660	3	4	2009	1	BHM	AL	TPA	FL		WN	WN	WN	1.00	G	460.00
200912901660	4	4	2009	1	TPA	FL	FLL	FL	X	WN	WN	WN	1.00	G	197.00
200912901661	1	4	2009	1	FLL	FL	BNA	TN		WN	WN	WN	1.00	G	793.00
200912901661	2	4	2009	1	BNA	TN	BUR	CA	X	WN	WN	WN	1.00	G	1790.00
200912901661	3	4	2009	1	BUR	CA	LAS	NV		WN	WN	WN	1.00	G	223.00

Figure 14: Excerpt from the DB1B Coupon database for 2009Q1. As for the Ticket database, the itinerary fare is nominal, in U.S.$, and distance is measured in miles. We will later take a closer look at the 3 highlighted itineraries.

airport it means that it is only a short-stop for the passenger.

- **Ticketing/Operating/Reporting carrier:** The ticketing system can sometimes be very complex, and the carrier which sells a flight is not necessarily the carrier which will operate it, nor will it necessarily be the carrier that reports the ticket to the Bureau of Transportation. Indeed, an itinerary containing several different coupons can be operated and sold by as many different carriers. In order to make our future analysis easier, we chose to retain only the operating carrier.

- **Passengers:** As for the Ticket database, we can see for a given coupon how many passengers bought a similar ticket. As a consequence, there is always a unique number of passengers for a given ItinID, in both databases. As for the Ticket database, this number can be very high, as the DOT prefers to combine similar tickets (same itineraries, same fares, same carriers,...)

in the quarter.

- **Fare class:** There are 6 different type of fare class defined by the DOT: split into restricted and unrestricted, there are first (unrestricted: code F; restricted: code G), business (respectively codes C and D) or coach class (Y or X) coupons. The DOT stresses that these classes are defined by carriers and can differ from one airline to another. Consequently any analysis of these fare classes should be interpreted with caution. For example, we can observe the distribution of the passengers in 2008 by fare class (Figure 15): a first remark is that this graph represents the fare class by coupons, and not by ticket, so the column axis represents number of passengers multiplied by the number of coupons bought by each passenger. Nevertheless, we can note that the proportion of passengers in the Business fare class only represents 1.13% of all coupons bought in 2008, whereas the coach class tickets represents more than 84% of all these coupons! This statistic shows in fact a characteristic of the U.S. domestic market, where the business fare class is not proposed by carriers as often as it is in Europe. The passenger discrimination is focused between 2 fare classes: the first class and coach class.

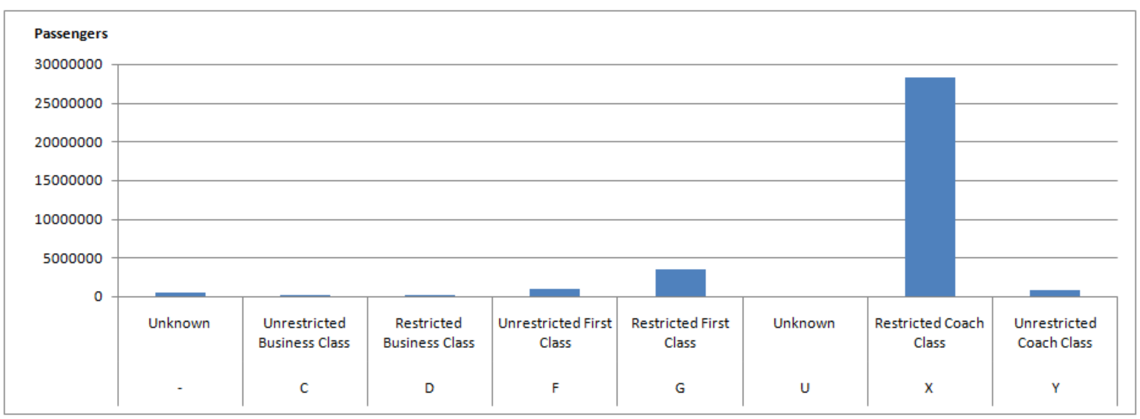


Figure 15: Fare class for coupons purchased in 2008. These fare class are defined by reporting carriers.

- **Distance:** We also know the length of each flight operated, in miles, which we could also compute from the exact location of the airports of origin and destination (latitudes, longitudes, giving the great-circle distance).

We selected 3 itineraries to illustrate this database (the 3 ItinId match the 3 examples taken above for the Ticket database). As a consequence, the data contained in the coupons with a specific ItinID should completely match the data for the corresponding ticket (that is indeed the case for the number of coupons in the itinerary, the period of time, the origin airport, the reporting carrier, the number of passengers and the distance). What can we learn from this database for our 3 examples?

- **ItinID 200912901650:** We learn from this ticket that the 3 passengers taking off from Fort Lauderdale in Florida went with their single flight to Nashville International Airport, in Tennessee, and that they chose to fly in the restricted first class.

- **ItinID 200912901658:** We also know that these passengers took a similar flight (as for the passengers with the Ticket 200912901650), but instead of staying in Tennessee, their second flight carried them to Birmingham-Shuttlesworth International Airport, in Alabama, 177 miles from Nashville. They flew in the unrestricted first class for the entire itinerary.

- **Itin 200912901660:** This last passenger also landed at Nashville airport with a similar flight. He stayed there for an unknown period, after which he took his return flight to Fort Lauderdale. This return flight unfolded with three legs, and this passenger landed successively at Birmingham-Shuttlesworth International Airport and at Tampa International Airport (in Florida) before reaching Fort Lauderdale.

To sum up, we can see that we are able to get considerable data about each ticket from the Coupon database. Nevertheless, the likely most important and interesting information about the itinerary, the fare of the ticket, is given only in the database Ticket (because we cannot of course give a fare for each of the coupons bought by the passengers).

3.1.3 Assumptions and inflation data

In order to obtain a reliable sample of data, and to avoid possible erroneous values in the Ticket and Coupon databases, we need to make some additional assumptions:

- **Passengers:** We observed that on some itineraries there were fewer than 90 passengers during a quarter. Given that the database represents 10% of all tickets sold in the U.S., there were less than 900 passengers for some itineraries during a quarter, which is equivalent to 75 passengers per week. There may be some flights operated with these 75 passengers per week; nevertheless, in order to delete what could be erroneous values in the database, **we decided to fix 90 passengers as a lower bound**.

- **Fares:** Another possible source of erroneous values in the Coupon and Ticket databases can be the fares. We make 2 additional assumptions:

  1. Inflation: The tickets reported by the carriers to the DOT in 1993 have been reported with the value of the dollar in 1993. Nevertheless, one dollar spent in 1993 does not offer the same purchasing power for the U.S. citizens as one dollar spent today. We consequently need to take into account the general rise in the level of prices (i.e. the inflation rate), in order to be able to compare the fare of a ticket bought in 1993 with the fare of a ticket bought in 2009. The usual method used to deal with this problem is the Consumer Price Index (CPI).[11] Below (Figure 16), we can see the monthly U.S. Consumer Price Index between 1993 and 2009 (available on the BLS website), which represents the increase in average prices over time. We note the sudden global decrease of prices in 2008, a clear sign of the recent economic and financial global crisis. We also observe that the Consumer Price Index is remarkably linear over the entire period.

     A simple regression (using least squares estimation) gives an estimated coefficient of

[11] See the U.S. Bureau of Labor Statistics website for more details on the CPI (`http://www.bls.gov/cpi/`).

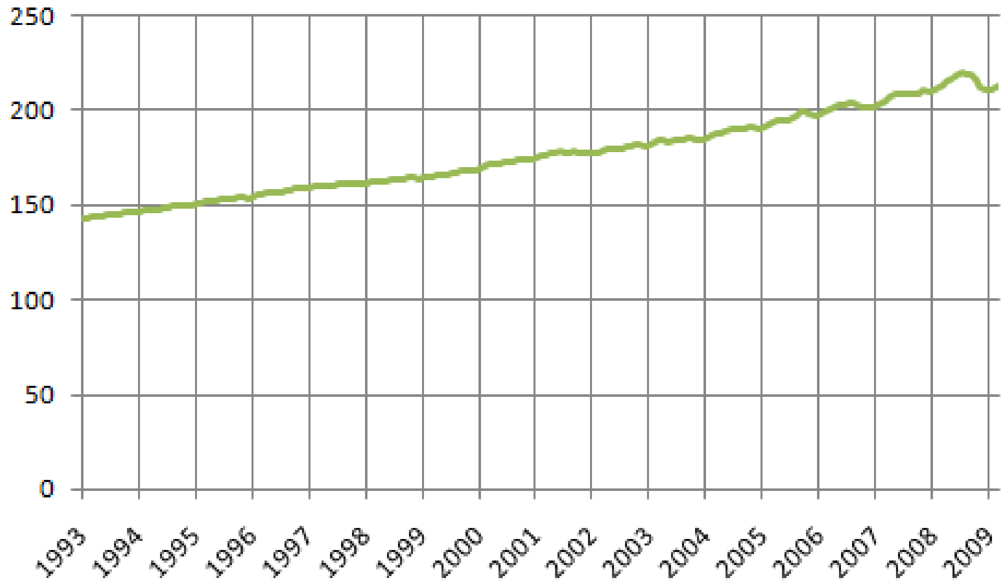


Figure 16: U.S. Consumer Price Index between 1993 and 2009.

1.126, meaning that there is an average increase of 1.126 in CPI per quarter, that is to say an average increase of 45 in the CPI index over 10 years. The econometric software EViews6 gives us a coefficient of R^2 (which represents the "quality" of fit for the regression) of 0.986, which is very high.

From these values, we can compute a "CPI deflator" value for each quarter, which will be a multiplying factor between the fares reported by the carriers in a given quarter and the real fares, taking into account the inflation of the prices between this quarter and today. The last quarter in our database is the first quarter in 2009 (afterwards this period will be written as 2009Q1), and our "original" value of the dollar will be the value for the quarter 2009Q1. The "CPI deflator" for this period (CPIdef(2009Q1)) is set at 1. The CPI deflator for the quarter 2008Q4 (CPIdef(2008Q4)) can be computed as the ratio between the quarterly average CPI of 2009Q1 (CPI(2009Q1)) and the quarterly average CPI of 2008Q4 (CPI(2008Q4)):

$$\text{CPIdef}(2008Q4) = \text{CPIdef}(2009Q1) \times \frac{\text{CPI}(2009Q1)}{\text{CPI}(2008Q4)}. \tag{1}$$

As a consequence, the CPI deflator for a given quarter t is:

$$\text{CPIdef}(t) = \text{CPIdef}(t+1) \times \frac{\text{CPI}(t+1)}{\text{CPI}(t)}. \tag{2}$$

Below (Figure 17), we see the Consumer Price Index deflators between 1993 and 2009 (dashed line), and the curve of the inflation of the quarterly average CPI (defined as the ratio between successive quarterly average CPI). As the CPI is almost a linear curve, therefore the CPI deflator curve is also approximately linear. On the other hand, the inflation of the quarterly average CPI is highly variable, especially for the past 10 years, reflecting the disturbance of the actual economy. We can note 3 major drops in quarterly inflation after 1993, leading to a quarterly deflation of the U.S. economy: the first one happened during 2001Q4, and is most likely to be a consequence of the September 11 attacks. The second one, which took place during 2006Q4, matches the beginning of the subprime mortgage financial crisis, with the end of the U.S. housing bubble. Finally, the third quarter of U.S. deflation, in 2008Q4, corresponded to the beginning of the current global economic crisis. We can check that the CPI deflator is a decreasing curve, with a minimum of 1 in the first quarter of 2009. The maximum value, 1.48, means as a consequence that we need to multiply all fares reported in the first quarter of 1993 by 1.48 to get the real value of fare which would be sold today (and a value we can therefore compare with the values of fares reported in the first quarter of 2009).

2. Zero-fare award coupons: Many U.S. airlines, and in our case each of the 7 carriers we are interested in, have put in place "frequent flyer programs": AAdvantage for American

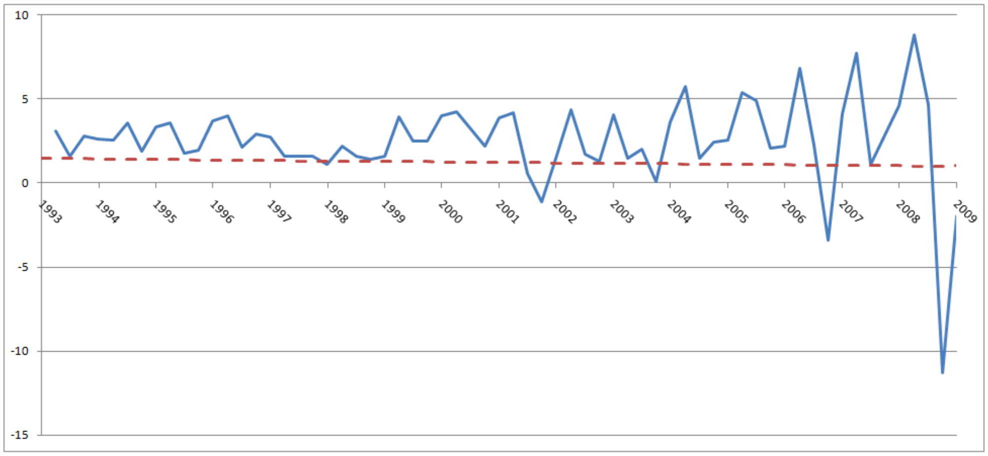


Figure 17: CPI deflators (dashed) and inflation estimates between 1993 and 2009.

Airlines, OnePass for Continental, SkyMiles for Delta and Northwest, Mileage Plus for United, Dividend Miles for US Airways and Rapid Rewards for Southwest. Due to these programs, some passengers can obtain tickets with zero-fare award coupons and pay nothing else but some taxes. Other passengers may also “fly for free”: the carrier employees. As we are not able to distinguish them from full-fare passengers in the databases, we must fix a lower bound for the fares in order to be able to correctly assess the real fares. This bound can also enable us to delete some other erroneous values. Which bound should we fix? We need to fix a bound that will not delete true low-fare tickets (offered by Southwest, for instance), but at the same time if this bound is too low, we may probably keep frequent flyer award coupon without any monetary payment but the air transport taxes.

3. Air transport taxes: All U.S. carriers are subject to taxes for each flight operated, most of which are reflected in the passenger’s ticket fare. What are the usual taxes paid by the passengers? The current aviation excise tax structure from the Federal

Aviation Authorities (FAA) is attached in the Appendix.[12] Nevertheless, the easiest way to find the additional taxes contained in a ticket is to look at the carriers' websites. Searching for the cheapest ticket sold by Southwest (on 29 November 2009) we found the following itinerary (Figure 18), from Kansas City to Milwaukee, at $59.60. According

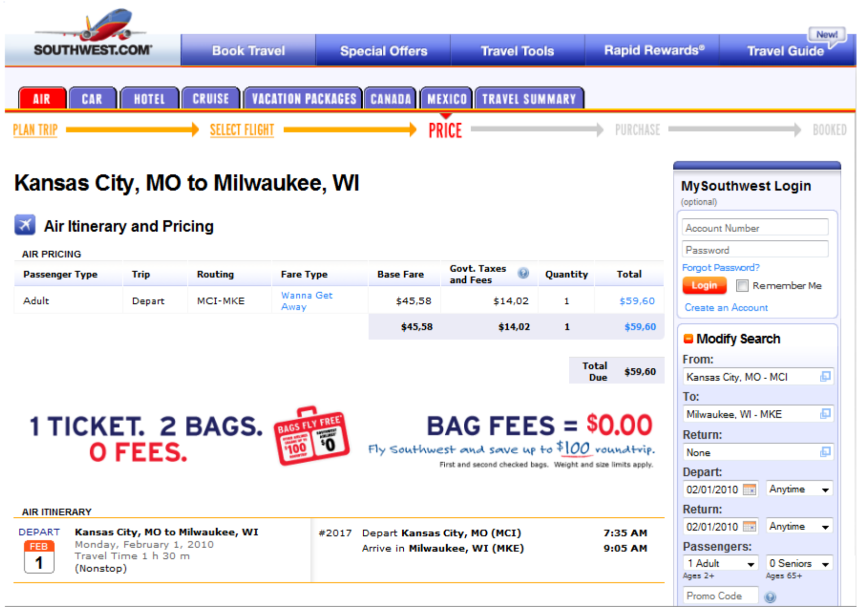


Figure 18: Illustrative Southwest ticket from Kansas City to Milwaukee.

to Southwest, about $14.02 on the $59.60 of the ticket are Government Taxes and Fees, made up with 7.5% (of the initial fare) for the Federal Excise Tax, accounting here for $3.42 (most of time, especially for the 6 other carriers, this tax is included in the initial fare proposed on the first booking screen), plus $3.60 of Federal segment fee (for each flight segment of the itinerary), plus $2.5 of "Government-imposed September 11th Security Fee" per segment (up to $5 for a one-way travel), and finally $4.5 of "Airport assessed Passenger Facility Charge" (PFC) per landing (up to $18 for a round trip, but

[12] A good report on usual aviation taxes and fees is also available at `http://assets.opencrs.com/rpts/RL33913_20080421.pdf` [8]

the PFC can be lower than \$4.5 for each segment, depending of the airport of origin on this segment). Checking the websites of the 6 other major carriers of interest, we observe that the additional taxes are the same as for Southwest.

As a consequence, we can assume that the minimal taxes paid for a one-way ticket (between big airports for which the PFC costs \$4.5) are \$10.6. On the other hand, the highest taxes, paid for a round trip with 4 successive segments, are up to \$42.4! It appears difficult to fix a reliable lower bound for fares, without losing some useful data on low-fare travels. In order to try to remove a large number of erroneous values, but without losing too much data, we decided to fix this **lower bound at \$20**.

4. Very high fares: A final source of error can be some high fare values. We saw in the database that some tickets are very expensive, and seem too expensive for domestic travel. As a consequence we decided to fix **a higher bound at \$10,000 for the fares**.

Once the lower and higher bounds for fares have been fixed, we can check in the database created with these assumptions whether these bounds are correctly fixed. We chose to gather all tickets bought by passengers in the same quarter for a round trip between the same airports and operated by the same carrier. We also create a fare for this round trip, fixed as the passenger-weighted average between all of the fares for this round trip during this quarter.

Do the assumptions for fares hold for this database?

We can see that our assumptions are quite realistic: we observe that the lowest real average fare in this database is \$60, and the highest fare is \$1801 (for round trips to the Island of Guam, in the western Pacific Ocean, i.e. more than 6000 miles from Los Angeles, compared to less than 3000 miles maximum for travel between the Pacific coast and the Atlantic cost of the U.S.!). These limits are easily contained in the interval fixed with the previous assumptions for fares.

4 Model construction and estimation

Here (Figure 19) we can see the quarterly average fares for round trips for the 7 carriers we are interested in, followed (Figure 20) by the quarterly number of passengers carried by these airlines for round trips between 1993 and 2009. We observe from these two figures a clear difference between

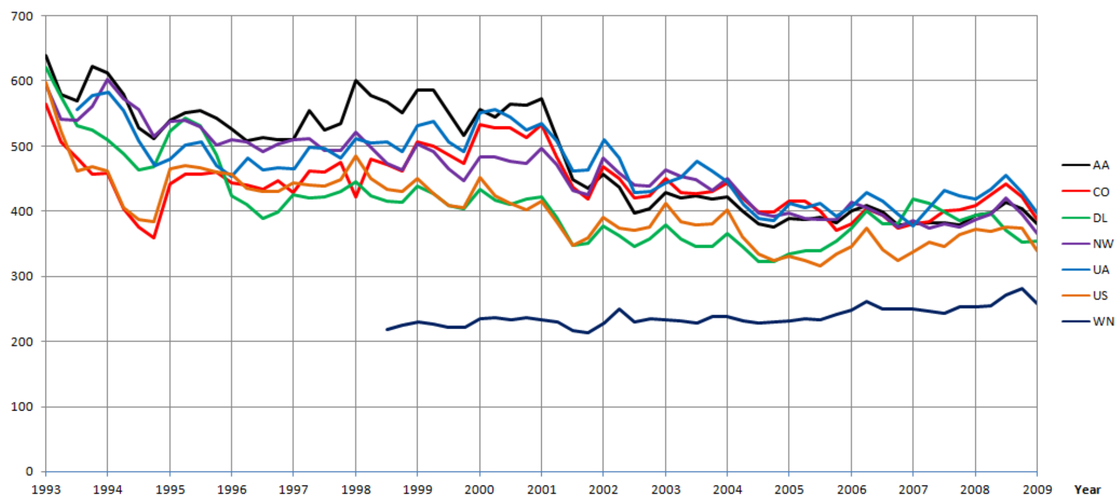


Figure 19: Quarterly average fares for round trips (unit: real 2009 US$). The data for the airline Southwest in the DB1B database, for round trips, is given by the DOT only from the third quarter of 1998. Before this quarter Southwest did not report round trips.

the low-cost model and the model for state airlines. Indeed, Southwest clearly has the lowest average fares, and is the largest airline in terms of passengers. In this section, we attempt to build rigorous forecasting models for the series displayed in these figures, from the simplest univariate time series models to more complicated regressions including both statistical and economic data.

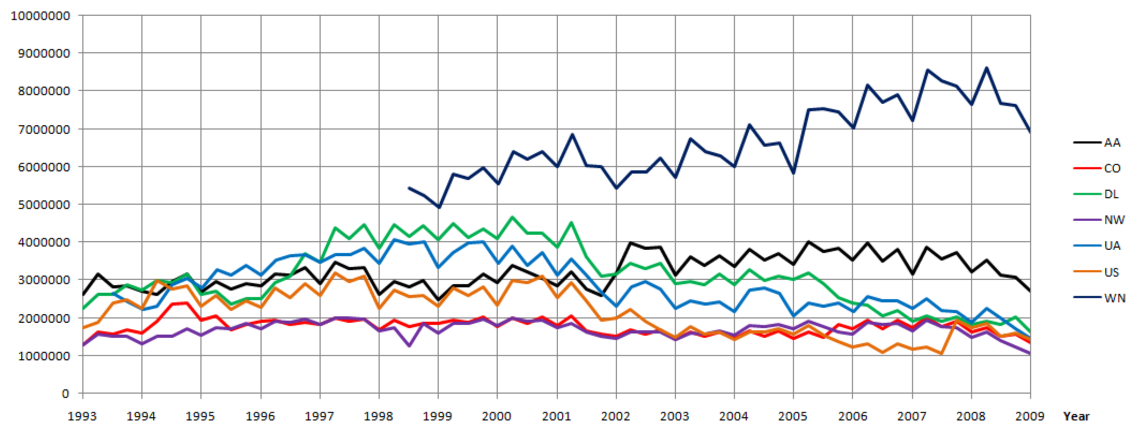


Figure 20: Number of passengers carried per quarter (round trips). As for the fares, data from Southwest is only available from the third quarter of 2008.

4.1 Fares

4.1.1 Simple linear regressions

One of the most straightforward ways to develop a model for the quarterly average fares, between 1993 and 2009, for a given carrier, is by a linear regression:

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times t, \tag{3}$$

where t is time. The limit of the fitted fares will be either $-\infty$, $+\infty$ or α_0 if $\alpha_1 = 0$, which is obviously very restrictive for any analysis and forecasting. In Table 1 we can see the estimated coefficients $\widehat{\alpha_0}$ and $\widehat{\alpha_1}$. First, we can observe that the Student's t-statistics are all greater (in absolute value) than 1.96, which means that we reject at the 95% level of confidence the hypothesis that a coefficient would be individually equal to 0.

It is interesting to see that the trend for all carriers is to decrease their fares, except for Southwest (this is apparent in the sign of the coefficient $\widehat{\alpha_1}$, which is positive only for the low-cost carrier).

Carriers	AA	CO	DL	NW	UA	US	WN
$\widehat{\alpha_0}$	602.86	479.13	493.94	559.32	540.30	476.91	198.28
$\widehat{\alpha_1}$	-3.769	-1.144	-2.567	-3.010	-2.123	-2.319	0.919
$\widehat{se}(\widehat{\alpha_0})$	9.678	9.761	10.166	4.816	9.241	7.767	5.017
$\widehat{se}(\widehat{\alpha_1})$	0.261	0.263	0.274	0.130	0.245	0.209	0.112
$t - stat(\widehat{\alpha_0})$	62.29	49.084	49.586	116.13	58.466	61.399	39.521
$t - stat(\widehat{\alpha_1})$	-14.45	-4.349	-9.365	-23.178	-8.656	-11.076	8.196
R^2	0.768	0.231	0.582	0.895	0.551	0.661	0.621

Table 1: The carriers are denoted by their IATA airline codes. "$se(\alpha_i)$" stands for the standard error of the coefficient α_i, and "$t - stat(\alpha_i)$" stands for the Student's t-statistic, which asymptotically follows a standard normal distribution. "R^2" is the coefficient of determination of the regression, which is a measure of the global fit of the model.

Nevertheless, in absolute value, Southwest is the most "stable" carrier in terms of variability of average fares.

Another interesting feature is the coefficient of determination of the regression, which appears to be very high for American Airlines and even moreso for Northwest Airlines. These coefficients can be interpreted as significant evidence for these two carriers to decrease their fares since 1993. It could be related to the difficult financial situation of the two airlines, since we know that Northwest airlines was forced to merge with Delta Airlines in 2008. For American Airlines, we can observe (see Figure 19) that the carrier was the "most expensive" after 2001. Following the successive political and economic crisis since 2001, a strategy to align the fares with the other airlines in a more competitive market may explain (more than the financial situation of the carrier) this decreasing trend for the fares of American Airlines since 1993.

4.1.2 ARMA models

Even if the linear model can explain a trend in the fares, it remains a very simple model and does not catch all the variations of fares since 1993. As a consequence, we could try a different approach to model the fare: the autoregressive moving average (ARMA) models. These models are often used for the analysis of statistical data. The main idea is to assume that the fare today will depend on the fares in previous periods (this is the autoregressive part of the model, also called the AR component), and on the shocks in these past fares altering the "average" value of the model (the moving average (MA) component of the model). An ARMA(p, q) model (combining AR(p) and MA(q) terms) can be written as follows:

$$\begin{aligned} \text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{Fare}(t-1) + \alpha_2 \times \text{Fare}(t-2) + ... + \alpha_p \times \text{Fare}(t-p) \\ + \varepsilon(t) + \beta_1 \times \varepsilon(t-1) + \beta_2 \times \varepsilon(t-2) + ... + \beta_q \times \varepsilon(t-q). \end{aligned} \tag{4}$$

We could imagine that the higher are p and q, the better the model could match reality. Nevertheless, by increasing the number of variables p and q, we decrease the number of available observations for our model and risk "overfitting". Given that our databases only have 65 quarterly observations from 1993 to 2009, we decided to limit p and q to 4, which correspond to one year of information for fares. These 4 past data also allow us to ignore for the moment possible seasonality effects on fares. As a consequence, we obtain for each carrier 25 new models (p and q =0,1,2,3,4).

We can also assume that the fares of these 7 carriers may also depend on the time period, and that there may be a trend in the fares to decrease or increase over time. In order to model this trend, we can add a variable of time in the regression of the fares. This new variable leads to 50 possible models of regression for the fares.

Moreover, another interesting variable to include in these regressions is a variable that could take into account what we can call a shock for the fares of the carrier (as a sudden and unexpected variation of the fares), caused by a major major external event such as an economic or political crisis, or an event such as the U.S. terrorist attacks of September 11, 2001. For example, looking at the curve of the fares for American Airlines (Figure 21), we can see that there is a visible gap between the average fares before and after the quarter 35, i.e. between July 2001 and September 2001 (included). This gap was the most important decrease of in average fares for American Airlines since 1993 (a drop of more than \$64 between the second and the third quarter of 2001), and was very likely caused by the terrorist attacks in New-York and the 3 days of non-activity imposed by the U.S. government after these events. Consequently, we can assume that we may need to take

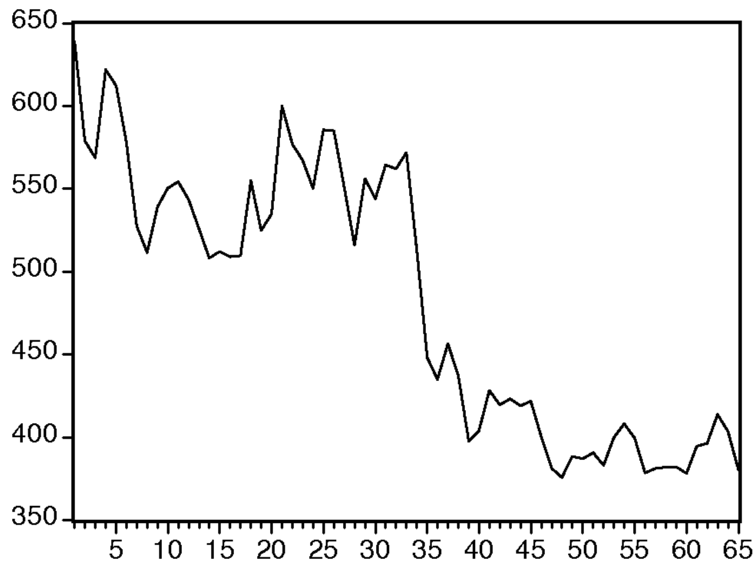


Figure 21: Quarterly average fares for American Airlines (\$ against time).

this shock into account in our model for fares. We chose to try to see the impact of this drop with the insertion of a dummy variable in our regression. How can we choose this dummy variable? In order to observe the change in the regression, we can insert a variable that takes value of 0 before the quarter 35, 0.22 for this quarter and 1 thereafter.[13] This dummy variable DummyD(t) can be

[13] The coefficient 0.22 was also used by Guzhva and Pagiavlas (2004) [9]. It represents the ratio 20/92 (20 days in the quarter 35 after 9/11 and 92 days total in this quarter).

expressed as follows :

$$\text{DummyD}(t) = 0 \text{ for } t < 35;\ 0.22 \text{ for } t = 35;\ 1 \text{ for } t \geq 35. \tag{5}$$

Thus we obtain 100 possible models for each airline.

4.1.3 Selecting the best ARMA model (theory)

So how can we choose the best among all these models? (that is, a reliable statistical and forecasting model which can also provide some interesting insights into the airlines' fare setting behaviour?) Many studies have been written on this subject, and methods have been proposed to find the best possible model for ARMA regressions. The classic approach is the Box-Jenkins methodology.[14] According to this procedure, one of the most important features to consider is the stationarity (stability) of a model. A model is said to be covariance-stationary if the mean, variance and covariance are finite and are invariant with respect to time. We can observe that some of the fares seem to decrease over time, and due to this "trend" we created models including a time variable. This variable can be a solution to "detrend" the regression and make this model trend-stationary. Nevertheless, even if we include a variable of time in the equation, the model can still be nonstationary. We can have a look, for example, at the following model AR(1) :

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{Fare}(t-1) + \varepsilon(t). \tag{6}$$

We can easily infer from this equation that if we know $\text{Fare}(t_0)$, we get for all $t \geq t_0$:

$$\text{Fare}(t) = \alpha_0 \times (1 + \alpha_1) + \alpha_1^2 \times \text{Fare}(t-2) + \alpha_1 \times \varepsilon(t-1) + \varepsilon(t), \tag{7}$$

[14] See for example `http://economics.uwaterloo.ca/2007-Winter/603notes3.pdf` for more information about this methodology, and Hamilton (1994) [10].

and so

$$\text{Fare}(t) = \alpha_0 \times (1 + \alpha_1 + \alpha_1^2 + ... + \alpha_1^{t-1}) + \alpha_1^t \times \text{Fare}(t_0)$$
$$+\alpha_1^{t-1} \times \varepsilon(t_0) + ... + \alpha_1 \times \varepsilon(t-1) + \varepsilon(t) \tag{8}$$

$$\Rightarrow \text{Fare}(t) = \alpha_0 \times \sum_{i=t_0}^{t-1}(\alpha_1^i) + \alpha_1^t \times \text{Fare}(t_0) + \sum_{i=t_0}^{t-1}(\alpha_1^i \times \varepsilon(t-i)). \tag{9}$$

What is the expected limiting behavior of Fare(t) as $t \to +\infty$? If we assume that $\varepsilon_t \sim D(0, \sigma^2)$, then $\lim_{t\to\infty} E[\sum_{i=t_0}^{t-1}(\alpha_1^i \times \varepsilon(t-i))] = 0$, where $E[.]$ is the expectation operator, $D(.)$ is some distribution and $\text{Fare}(t_0) = 0$ is assumed. and we can observe that:

- if $|\alpha_1| < 1$: $\lim_{t\to\infty} E[\text{Fare}(t)] = \frac{\alpha_0}{1-\alpha_1}$
- if $|\alpha_1| = 1$: $\lim_{t\to\infty} \text{Fare}(t) = \infty$ if $\alpha_0 \neq 0$, $\lim_{t\to\infty} \text{Fare}(t) = 0$ if $\alpha_0 = 0$
- if $|\alpha_1| > 1$: $\lim_{t\to\infty} \text{Fare}(t) = \infty$,

and so the AR(1) model is not explosive if and only if $|\alpha_1| < 1$. In fact, we can find a general rule for all ARMA models, which determines a condition in the α terms to get a finite limit for a model and because of that the stationarity (stability) of this model. To establish this condition, we need to rewrite the ARMA(p,q) models as:

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{Fare}(t-1) + \alpha_2 \times \text{Fare}(t-2) + ... + \alpha_p \times \text{Fare}(t-p)$$
$$+\beta_1 \times \varepsilon(t-1) + \beta_2 \times \varepsilon(t-2) + ... + \beta_q \times \varepsilon(t-q). \tag{10}$$

Let L^i be the lag operator, which transforms Fare(t) into Fare($t-i$). Then we can write the model ARMA(p,q) as follows:

$$\text{Fare}(t) = \sum_{i=0}^{p}(\alpha_i \times \text{L}^p) \times \text{Fare}(t) + \sum_{i=1}^{q}(\beta_i \times \text{L}^q) \times \varepsilon(t) \tag{11}$$

or

$$(1-\sum_{i=1}^{p}(\alpha_i \times \mathrm{L}^p)) \times \mathrm{Fare}(t) = \alpha_0 + \sum_{i=1}^{q}(\beta_i \times \mathrm{L}^q) \times \varepsilon(t), \tag{12}$$

and if we denote

$$A(L) := 1-\sum_{i=1}^{p}(\alpha_i \times L^p) \tag{13}$$

$$B(L) := \sum_{i=1}^{q}(\beta_i \times L^q) \tag{14}$$

then

$$A(L) \times \mathrm{Fare}(t) = \alpha_0 + B(L) \times \varepsilon(t). \tag{15}$$

The stability condition for the ARMA(p,q) model is the following: under the assumption that $\varepsilon(t) \sim D(0, \sigma^2)$ (meaning that ε follows an independent, identically distributed, or White Noise distribution), a model is stationary if all the roots of the real polynomial $\alpha_Z = 1 - \sum_{i=1}^{p}(\alpha_i \times Z^p)$ are greater than 1 in absolute value (or in norm for complex roots).

If this stability condition holds, then we can invert the polynomial α_Z (which can be done thanks to the convolution formula), and in the same way we get the inverse of the polynomial $A(L)$, $A(L)^{-1}$, and thus we can "solve" the equation Fare(t):

$$\mathrm{Fare}(t) = \alpha_0 \times A(L)^{-1} + A(L)^{-1} \times B(L) \times \varepsilon(t). \tag{16}$$

As α_0 is a constant, in fact we can prove that $\alpha_0 \times A(L)^{-1} = \alpha_0 \times A(1)^{-1}$, so the limit of the fares is:

$$\lim_{t\to\infty} \mathrm{Fare}(t) = \alpha_0 \times A(1)^{-1}. \tag{17}$$

The model is therefore stationary, as we know that the limit of the fares is finite.

EViews6 can provide us with the "inverted" roots of the AR polynomial $A(L)$, in the estimation output of an ARMA model. These roots are in fact the inverted roots of the polynomial α_Z. Indeed, we can prove that the stability conditions is equivalent to having these inverted roots lower in absolute value (or in norm) than 1. So we can check, for each model, whether the stability condition holds. However, we can observe that in some cases the roots are below (but very close to) 1. This proximity to the limit value of stability can be a problem as it makes it very difficult to find $A(L)^{-1}$, and such values can be a sign that the model is not as stable as we would wish. Some tests can allow us to solve this problem and determine a critical value under which we may expect to have stationarity with a certain amount of confidence. One of the tests we can perform with EViews6 is the Augmented Dickey-Fuller (ADF) test, which checks whether a unit root is present or not in the model. The ADF test uses the following regression, with the difference operator notation $\Delta\text{Fare}(t-p) := \text{Fare}(t-p) - \text{Fare}(t-p-1)$:

$$\begin{aligned}\Delta\text{Fare}(t) = \alpha_0 \times \text{Fare}(t-1) + \alpha_1 \times \Delta\text{Fare}(t-1) + \alpha_2 \times \Delta\text{Fare}(t-2) + ...\\ +\alpha_p \times \Delta\text{Fare}(t-p) + \varepsilon(t)\end{aligned}. \tag{18}$$

We can add to this equation a constant term, and a time trend, to check the stationarity of these models too. Then what we want to test is whether $\alpha_0 = 0$. The other $\Delta\text{Fare}(t-i)$ terms behind are used to avoid possible autocorrelation effects of the residuals in the regression. This regression allow us to compute the t-statistic for the coefficient α_0. Then we reject the hypothesis of the null for this coefficient (and so we reject the nonstationarity hypothesis for the model) if this computed t-statistic is greater than the critical value at a certain level of confidence.

If we find, thanks to the ADF test, that a model is still nonstationary, then there is a common tool used in econometrics, which is to apply an initial differencing step to the model, and to transform the ARMA model into an autoregressive integrated moving average (ARIMA) model. In this ARIMA model we consequently do not estimate $\text{Fare}(t)$ but $\Delta\text{Fare}(t-1)$, which may "stationarize" the regression.[15] We may need to difference the data several times (using ΔFare($t-2$) or ΔFare($t-3$) as the dependent variable for the regression) to get stationarity, but this method can sometimes create additional problems if the differencing is taken "too far", and so we will try to stay at the first-order of differencing. We can also evaluate, with the Dickey-Fuller test, whether the ARIMA models have unit roots or not, and thus we can check the stationarity of these models too.

Once we obtain a stationary model, we also need to perform other tests on the regression output to select the best one. We first need to check that all the coefficients of the regression are significant (considering the Student's t-statistics of all coefficients). Then we need to check that the residuals have nice distributional properties (checking that they follow a White Noise distribution), and that they are not autocorrelated (the Durbin-Watson statistic must be close to 2).[16] The autocorrelation of the residuals can also be checked using the correlograms and the Ljung-Box Q-statistic. Indeed, we can observe on the correlograms the (partial) autocorrelation functions of the residuals, which represent the dependence between a value of the fares at the time period t (Fare(t)), and the value of these fares s periods before t (Fare($t-s$)). If the value of this dependence is too high, it means that there is a strong dependence between the fares at these two periods of time, and therefore our model may have missed an important correlation in the regression. Thus we need to check whether there is a spike in the correlograms, and if such a spike exists what is the probability of dependence between the two fares (thanks to the Q-statistic).

Finally, how can we choose, among these good models, the best one? The Schwarz Information Criterion (SIC) can be a solution to select the model which best fits the model. The SIC is asymptotically consistent for ARMA(p,q) models, i.e. given a "large enough" sample and a set of p and q that includes the true p,q, the SIC will select the correct lag-order. The criterion is given

[15] See `http://www.duke.edu/~rnau/411arim.htm#les` for details.

[16] See `http://en.wikipedia.org/wiki/Durbin-Watson_statistic` for further information on the Durbin-Watson statistic.

by

$$\text{SIC} = \frac{-2 * \ln L}{T} + \frac{k * \ln T}{T}, \tag{19}$$

where L is the maximized log-likelihood (estimated), T is the effective sample size, and k is the number of estimated parameters. The best ARMA model is chosen by minimizing the SIC across various ARMA(p,q) models, e.g. $p, q = 0, 1, 2, 3, 4$.

4.1.4 Application of ARMA models to 7 U.S. carriers

If we run a first series of regressions for the seven carriers, and keep the best model among these regressions according the Schwarz Information Criterion, we get the following results:

- **American:** The best ARMA model for American Airlines seems to be an ARMA(3,0), that is to say an AR(3). It would mean that the best ARMA regression for the airline would be:

  $$\text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{Fare}(t-1) + \alpha_2 \times \text{Fare}(t-2) + \alpha_3 \times \text{Fare}(t-3) + \varepsilon(t). \tag{20}$$

  The Student's t-statistics for the coefficients of the regression are greater (in absolute value) than 1.96, so they are individually significant. The coefficient R^2 is equal to 0.918, which is very high. The Durbin-Watson statistic is equal to 2.015, close to 2, meaning that the residuals of the regression do not appear to be autocorrelated. The Jarque Bera statistic is quite low (1.390), so we can assume that these residuals are approximately normally distributed.

  However, the very high coefficient of determination R^2 is a sign that there may be a problem hidden behind these apparently good results. Looking at the inverse roots of the AR polynomial, we see that one root is equal to 0.968, which is very close to 1. There may consequently be a problem of nonstationarity for this model (the model is unstable). This hypothesis is confirmed by the unit root test.

  Thus, we can try to difference this equation to remove this nonstationarity: the ARMA model

ARMA(3,0) with first-order differencing leads to an ARIMA(3,1,0):

$$\begin{aligned} \text{Fare}(t) - \text{Fare}(t-1) &= \alpha_0 + \alpha_1 \times (\text{Fare}(t-1) - \text{Fare}(t-2)) \\ +\alpha_2 \times (\text{Fare}(t-2) - \text{Fare}(t-3)) &+ \alpha_3 \times (\text{Fare}(t-3) - \text{Fare}(t-4)) + \varepsilon(t). \end{aligned} \tag{21}$$

In order to get the best differenced models for our regressions, we consequently keep the ARIMA model minimizing the Schwarz criterion.

For American Airlines, the result of this differencing appears to be the ARIMA(2,1,0) model (i.e. ARMA(2,0) with first-order differencing). Nevertheless, we can observe that, even though the inverted roots of the AR polynomials are here far from 1 in absolute value, the other statistics are worsened by the differencing. The R^2 coefficient is only equal to 0.11 (meaning that the model would explain roughly 11% of the data for fares), and, for one coefficient, we do not reject at the 95% level of confidence the hypothesis that it is equal to 0. These results tend to say we may have missed an important characteristic in our model for this airline.

We can first try to see if there is a change in the trend of the airline for its average fares after September 11, 2001. We run the following regression:

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{D}(t) + \alpha_2 \times t + \alpha_3 \times \text{D}(t) \times t + \varepsilon(t). \tag{22}$$

From the results of this regression (Figure 22), we can see that the intercept dummy variable appears to be very important and reveals a drop of about \$100 in the average fares for the airline during the September 11 attacks. We can also observe that there is a trend in the fares (we reject the hypothesis of the null for the time trend variable), but this trend does not seem to change after 2001, as the variation of this coefficient is very low (\$0.37, with a standard error of 0.75, hence we do not reject the null hypothesis for this coefficient at the

95% level of confidence). So maybe we could include this dummy variable, with a trend, in a

Dependent Variable: FARE1
Method: Least Squares
Date: 12/30/09 Time: 17:24
Sample: 1 65
Included observations: 65

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	569.8331	9.291963	61.32537	0.0000
DUMMY	-99.50148	30.86451	-3.223815	0.0020
@TREND	-1.023797	0.475261	-2.154177	0.0352
@TREND*DUMMY	-0.372762	0.749828	-0.497130	0.6209

R-squared	0.887526	Mean dependent var	482.2520
Adjusted R-squared	0.881995	S.D. dependent var	81.31382
S.E. of regression	27.93281	Akaike info criterion	9.557045
Sum squared resid	47594.76	Schwarz criterion	9.690853
Log likelihood	-306.6040	F-statistic	160.4499
Durbin-Watson stat	0.904774	Prob(F-statistic)	0.000000

Figure 22: American Airlines - regression on quarterly average fares: @trend stands for the regression of the results over time, and dummy stands for the dummy variable previously presented.

ARMA model, and get the best ARMA model according to the Schwarz criterion. Here is the result of this new regression: EViews6 indicates that the best model with these requirements is an ARMA(0,2) (Figure 23), which can be written:

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times t + \alpha_2 \times \text{D}(t) + \varepsilon(t) + \beta_1 \times \varepsilon(t-1) + \beta_2 \times \varepsilon(t-2). \tag{23}$$

Indeed we can see that this model is much better than the previous ones: all coefficients are individually significant, and jointly significant (the F-statistic is very high). The inverted MA roots are lower than 1 in absolute value, so the model is invertible, and also stationary (confirmed by the unit root test). The Jarque-Bera statistic is equal to 6.68 (not present in the Figure), so we reject the normality of the residuals. The Durbin-Watson statistic is equal to 2.176 so there is no autocorrelation either in these residuals, and there are no significant spikes in the autocorrelation function (ACF) or the partial autocorrelation function (PACF), so the residuals of this regression are similar to White Noise, and there seem to be no other significant patterns left in the data. To sum up, we can notice that the quarterly average

Dependent Variable: FARE1
Method: Least Squares
Date: 12/30/09 Time: 16:44
Sample: 1 65
Included observations: 65
Convergence achieved after 11 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	574.6830	13.56196	42.37462	0.0000
@TREND	-1.640269	0.589527	-2.782348	0.0072
DUMMY	-88.83975	21.54177	-4.124070	0.0001
MA(1)	0.882115	0.095828	9.205166	0.0000
MA(2)	0.385701	0.093419	4.128723	0.0001

R-squared	0.934305	Mean dependent var	482.2520
Adjusted R-squared	0.929926	S.D. dependent var	81.31382
S.E. of regression	21.52502	Akaike info criterion	9.050112
Sum squared resid	27799.59	Schwarz criterion	9.217373
Log likelihood	-289.1287	F-statistic	213.3293
Durbin-Watson stat	2.176079	Prob(F-statistic)	0.000000
Inverted MA Roots	-.44-.44i	-.44+.44i	

Figure 23: American Airlines - final regression on quarterly average fares.

fares for American Airlines (for round trips) fell substantially after the September 11 terrorist attacks, by about \$90 on average. There is also a trend for these fares to decrease over time, by about \$1.64 per quarter.

Remark: these regressions were executed with a lower bound in the database. Indeed, we decided to keep only the flights with more than 90 passengers. Nevertheless, we observe that the results are the same if we consider the non-filtered data, and the model ARMA(0,2), with a dummy variable and a trend in the fares, remains the best model for this carrier. We can also observe that the Schwarz criterion is lower for the non-filtered data, so the use of this bound may not be that relevant for American Airlines. Thus we choose to work with the regression output of the non-filtered data (Figure 24).

Dependent Variable: FARE1
Method: Least Squares
Date: 01/08/10 Time: 16:38
Sample: 1 65
Included observations: 65
Convergence achieved after 12 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	580.3625	13.18745	44.00870	0.0000
DUMMY	-87.85106	21.04902	-4.173642	0.0001
@TREND	-1.650983	0.574834	-2.872104	0.0056
MA(1)	0.860514	0.094353	9.120151	0.0000
MA(2)	0.370623	0.092009	4.028100	0.0002

R-squared	0.934410	Mean dependent var	488.0287
Adjusted R-squared	0.930037	S.D. dependent var	80.47577
S.E. of regression	21.28624	Akaike info criterion	9.027802
Sum squared resid	27186.24	Schwarz criterion	9.195063
Log likelihood	-288.4036	F-statistic	213.6928
Durbin-Watson stat	2.220186	Prob(F-statistic)	0.000000
Inverted MA Roots	-.43+.43i	-.43-.43i	

Figure 24: American Airlines - final regression on quarterly average fares (non filtered data).

- **Continental:** We apply the same methodology as for American Airlines: the same initial process used for the quarterly average fares for Continental leads to the ARMA model ARMA(1,0), as the best ARMA model we can get according to the Schwarz criterion (see Figure 25). This model seems to be quite good, as the AR root is equal to 0.77, so the model is stationary. All the coefficients are individually significant, and jointly significant according to the F-statistic. Nevertheless, we can observe that the coefficient of determination is quite low ($R^2 = 0.66$), and more importantly, the Durbin-Watson statistic is also quite low (1.83), which could mean that there is a positive autocorrelation in the residuals. This is confirmed by the correlograms of the regression, in which we can see a spike of autocorrelation in the PACF function for the fourth lag (see Figure 26). As a consequence, we may need to reconsider our regression. This autocorrelation may be caused by a seasonality in the average fares of Continental, but it could also be caused by a shock to the fares.

Dependent Variable: FARE2
Method: Least Squares
Date: 01/18/10 Time: 16:36
Sample (adjusted): 2 65
Included observations: 64 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	434.2499	12.84776	33.79967	0.0000
AR(1)	0.756818	0.073189	10.34066	0.0000

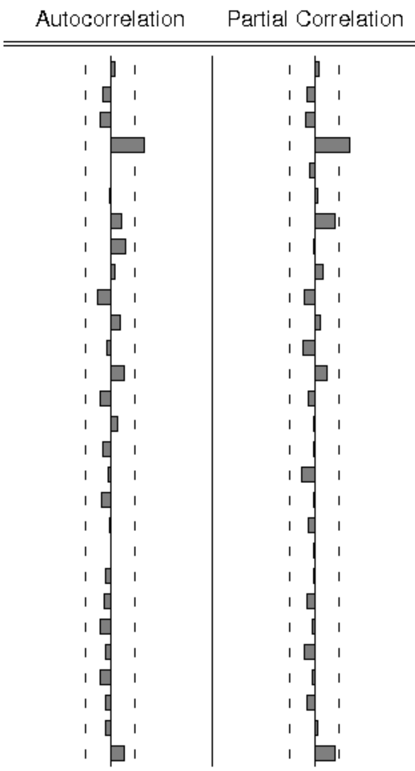
R-squared	0.632982	Mean dependent var	442.7685
Adjusted R-squared	0.627063	S.D. dependent var	39.48063
S.E. of regression	24.11026	Akaike info criterion	9.233903
Sum squared resid	36040.89	Schwarz criterion	9.301369
Log likelihood	-293.4849	F-statistic	106.9292
Durbin-Watson stat	1.847370	Prob(F-statistic)	0.000000
Inverted AR Roots	.76		

Figure 25: Continental Airlines - regression on quarterly average fares.

Indeed, if we look at the figure of the fares (Figure 27), we observe 2 major breaks in the fares, a sudden drop in the fares between the periods 5 and 8 (from the second quarter of 1994 to the first quarter of 1995), and a second one in period 35. This first drop is probably related to the period of troubles faced by the carrier in 1994 and the fact that Continental was only just emerging from bankruptcy that year. In the same period, we can observe a major increase in the number of passengers carried by the airline (Figure 52), so the drop in fares may be part of the strategy of the airline at the time. It could consequently be a mistake to consider that this drop in fares is caused by an event independent from the airline (as the terrorist attacks of 2001). The second drop happened during the terrorist attacks in 2001, so we could try to include a dummy variable in our regression as we did for American Airlines.

In fact, it turns out that if we include the same dummy variable used for American Airlines, we get a totally different model, a ARMA(0,2) with a dummy variable (see Figure 29). We

Date: 01/18/10 Time: 16:48
Sample: 2 65
Included observations: 64
Q-statistic probabilities adjusted for 1 ARMA term(s)



	AC	PAC	Q-Stat	Prob
1	0.044	0.044	0.1271	
2	-0.076	-0.078	0.5160	0.473
3	-0.099	-0.093	1.1963	0.550
4	0.350	0.358	9.8149	0.020
5	0.005	-0.057	9.8167	0.044
6	-0.005	0.037	9.8189	0.081
7	0.108	0.206	10.684	0.099
8	0.148	-0.011	12.327	0.090
9	0.044	0.084	12.476	0.131
10	-0.124	-0.112	13.681	0.134
11	0.104	0.052	14.542	0.150
12	-0.037	-0.117	14.654	0.199
13	0.142	0.117	16.317	0.177
14	-0.099	-0.068	17.146	0.193
15	0.072	-0.009	17.589	0.226
16	-0.074	-0.009	18.075	0.259
17	-0.022	-0.137	18.117	0.317
18	-0.090	-0.012	18.861	0.337
19	-0.012	-0.063	18.874	0.400
20	0.005	-0.008	18.877	0.465
21	-0.047	-0.011	19.092	0.516
22	-0.067	-0.078	19.541	0.550
23	-0.111	-0.028	20.812	0.532
24	-0.051	-0.106	21.085	0.576
25	-0.104	-0.020	22.250	0.564
26	-0.048	-0.077	22.511	0.606
27	-0.046	0.028	22.749	0.647
28	0.145	0.208	25.219	0.562

Figure 26: Continental Airlines - residual correlograms of the regression in Figure 25.

can observe that the model is better according to the Schwarz Information Criterion (lower for this second regression). The R^2 is also higher than for the first regression. All the coefficients are still significant, the model is stationary, and the Durbin-Watson statistic exceeds 2, which could mean that there is negative autocorrelation in the residuals. Checking the correlograms (Figure 30), we observe that there is no remaining autocorrelation in the residuals at the 95% level of confidence. We can observe that including a trend in this model does not improve the fit. The t-statistic for this trend (the trend coefficient being equal to 0.45 according to

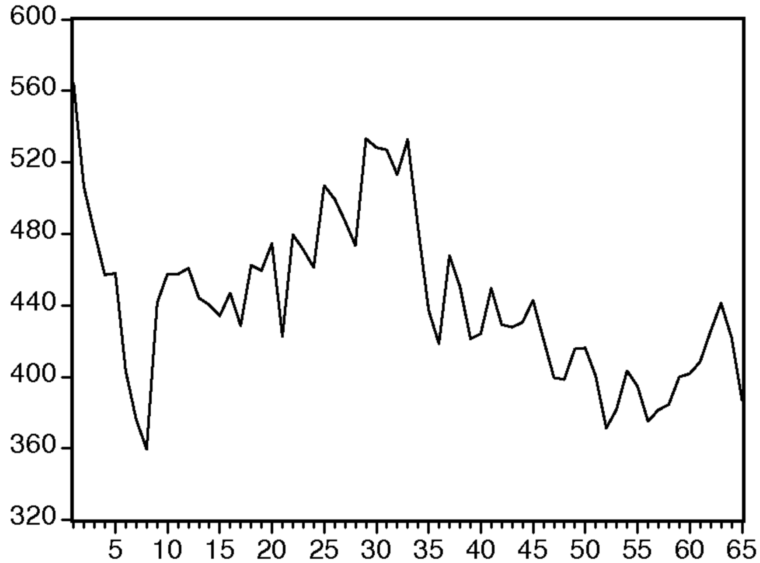


Figure 27: Quarterly average fares for Continental Airlines ($ against time).

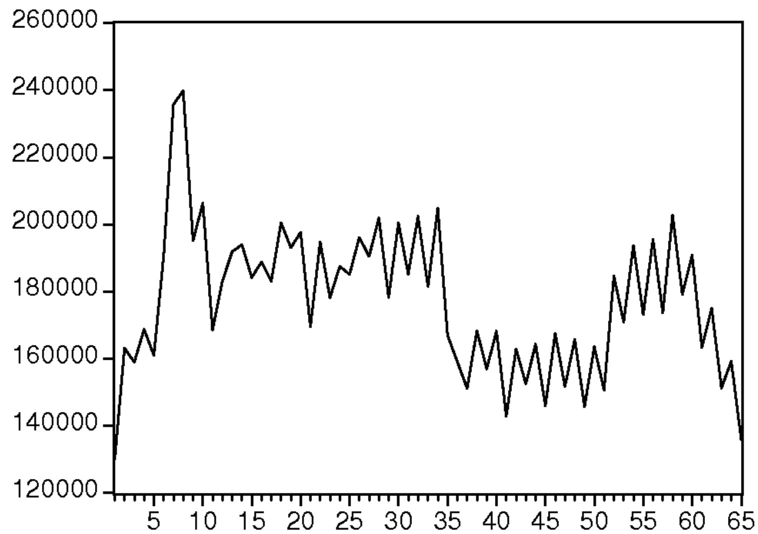


Figure 28: Number of passengers carried by Continental Airlines per quarter.

Dependent Variable: FARE2
Method: Least Squares
Date: 01/18/10 Time: 16:59
Sample (adjusted): 1 65
Included observations: 65 after adjustments
Convergence achieved after 10 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	466.3134	9.208430	50.63983	0.0000
DUMMY	-51.89094	12.92411	-4.015049	0.0002
MA(1)	0.926233	0.076501	12.10751	0.0000
MA(2)	0.548353	0.083744	6.547973	0.0000

R-squared	0.734066	Mean dependent var	444.5904
Adjusted R-squared	0.720987	S.D. dependent var	41.83435
S.E. of regression	22.09761	Akaike info criterion	9.088380
Sum squared resid	29786.57	Schwarz criterion	9.222188
Log likelihood	-291.3723	F-statistic	56.12665
Durbin-Watson stat	2.202367	Prob(F-statistic)	0.000000
Inverted MA Roots	-.46+.58i	-.46-.58i	

Figure 29: Continental Airlines - regression on quarterly average fares.

EViews6) is equal to 0.66, so we do not reject the hypothesis that this coefficient is null at the 95% level of confidence.

Moreover, if we use the non-filtered data (without the lower bound for the flights with more than 90 passengers), we observe that the best ARMA model is also the ARMA(0,2) model with a dummy variable. The coefficient of determination R^2 is equal to 0.73, all the coefficients are individually significant, and the model is stationary. The Durbin-Watson statistic is also higher than 2. Moreover the SIC criterion is lower for this regression than for the regression with the filtered data, and so we can infer that this model better fits the data.

In order to improve our model, we could then try to catch possible seasonality in the fares as seen in the first regression (with autocorrelation at the fourth lag) for the model ARMA(1,0). If we include an autoregressive term of order 4 in our regression, which would "capture" the autocorrelation of the fourth lag, we can see that this solution does not improve the model.

Date: 01/18/10 Time: 17:03
Sample: 1 65
Included observations: 65
Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.111	-0.111	0.8360	
		2	0.022	0.010	0.8697	
		3	0.066	0.070	1.1720	0.279
		4	0.249	0.269	5.6080	0.061
		5	-0.050	0.010	5.7907	0.122
		6	0.041	0.022	5.9135	0.206
		7	0.156	0.134	7.7483	0.171
		8	0.101	0.083	8.5337	0.202
		9	0.039	0.072	8.6493	0.279
		10	-0.105	-0.147	9.5249	0.300
		11	0.154	0.041	11.426	0.248
		12	-0.176	-0.223	13.960	0.175
		13	0.162	0.119	16.160	0.135
		14	-0.119	-0.093	17.370	0.136
		15	-0.015	-0.108	17.389	0.182
		16	-0.083	-0.049	17.998	0.207
		17	-0.036	-0.120	18.113	0.257
		18	-0.085	-0.020	18.782	0.280
		19	-0.040	-0.002	18.937	0.332
		20	-0.046	-0.028	19.140	0.383
		21	-0.097	-0.037	20.075	0.390
		22	-0.047	-0.062	20.298	0.439
		23	-0.143	-0.030	22.433	0.375
		24	-0.031	-0.074	22.533	0.428
		25	-0.100	0.025	23.631	0.424
		26	-0.033	-0.050	23.750	0.476
		27	-0.014	0.068	23.772	0.533
		28	0.075	0.168	24.440	0.551

Figure 30: Continental Airlines - residual correlograms of the regression in Figure 29.

We get the model ARMA(0,2) with a AR(4) term and our dummy variable (Figure 31), which can be written:

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{Dummy}(t) + \alpha_2 \times \text{Fare}(t-4) + \varepsilon + \beta_1 \times \varepsilon(t-1) + \beta_2 \times \varepsilon(t-2). \quad (24)$$

There is no problem of autocorrelation for this model. The coefficient for the fourth lag is significant at the 97% level of confidence. All the other coefficients also remain individually significant. The coefficient of determination R^2 for this regression is equal to 0.71. It was

Dependent Variable: FARE2
Method: Least Squares
Date: 01/10/10 Time: 15:13
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 10 iterations
Backcast: 3 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	463.1689	12.28482	37.70254	0.0000
DUMMY	-49.20115	15.49425	-3.175446	0.0024
AR(4)	0.314990	0.132395	2.379172	0.0208
MA(1)	0.799716	0.125517	6.371387	0.0000
MA(2)	0.431340	0.123622	3.489191	0.0010

R-squared	0.712162	Mean dependent var	440.5456
Adjusted R-squared	0.691602	S.D. dependent var	39.03847
S.E. of regression	21.67947	Akaike info criterion	9.069021
Sum squared resid	26319.97	Schwarz criterion	9.242043
Log likelihood	-271.6051	F-statistic	34.63844
Durbin-Watson stat	1.956453	Prob(F-statistic)	0.000000

Inverted AR Roots	.75	.00+.75i	-.00-.75i	-.75
Inverted MA Roots	-.40+.52i	-.40-.52i		

Figure 31: Continental Airlines - regression on quarterly average fares.

about 0.73 for the previous regression, so we do not lose much significance with this new model, but as for the filtered data, the SIC criterion is higher with this additional term than for the model in Figure 29, so we decide to keep the previous model (ARMA(0,2) with a dummy variable).

- **Delta:** For Delta Airlines, the same methodology leads to an ARMA(3,2) model with a trend variable for the filtered data (see Figure 32), and the model ARMA(4,3) for the whole database. Nevertheless, these models are very complex, and show possible problems of non-stationarity in the AR roots (very close to 1), even if all the coefficients are significant. If we include a dummy variable in the regression, to catch possible external effects from the September 11, 2001, we observe that the model looks simpler, as we obtain a ARMA(0,3) with a dummy variable. All the coefficients remain significant, the R^2 statistic is higher than

Dependent Variable: FARE3
Method: Least Squares
Date: 01/18/10 Time: 17:35
Sample (adjusted): 4 65
Included observations: 62 after adjustments
Convergence achieved after 11 iterations
Backcast: 2 3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	379.8326	26.79139	14.17741	0.0000
AR(1)	0.836018	0.107980	7.742305	0.0000
AR(2)	-0.811370	0.131557	-6.167449	0.0000
AR(3)	0.769838	0.099230	7.758092	0.0000
MA(1)	0.426887	0.109506	3.898309	0.0003
MA(2)	0.849721	0.097860	8.683040	0.0000

R-squared	0.905418	Mean dependent var	403.8854
Adjusted R-squared	0.896973	S.D. dependent var	52.89033
S.E. of regression	16.97665	Akaike info criterion	8.593320
Sum squared resid	16139.57	Schwarz criterion	8.799172
Log likelihood	-260.3929	F-statistic	107.2155
Durbin-Watson stat	1.898956	Prob(F-statistic)	0.000000

Inverted AR Roots	.89	-.03+.93i	-.03-.93i
Inverted MA Roots	-.21+.90i	-.21-.90i	

Figure 32: Delta Airlines - regression on quarterly average fares.

for the first regression, but the SIC criterion has been increased. Looking at the fares, we can observe that including a dummy variable in the regression at period 35 seems to be a good idea as we can observe a drop in the fares at this period (Figure 33). However, could this drop reflect a more global decreasing trend in the fares of the airline? In fact we can observe that, if we include a trend in our model, we get quite a good model with a low SIC criterion, for the filtered data. For the whole database, this variable seems to add complexity to the model rather than improving it (we get a ARMA(4,1) model, with autocorrelation problems). For the filtered data, we get a ARMA(1,1) with the trend variable (see Figure 34). 34). This model is stationary, and all the coefficients seem to be significant in the regression (looking at the Student's t-statistics). The trend coefficient is equal to -1.83. The coefficient of determination for this regression is equal to 0.91, and the SIC is equal to 8.83,

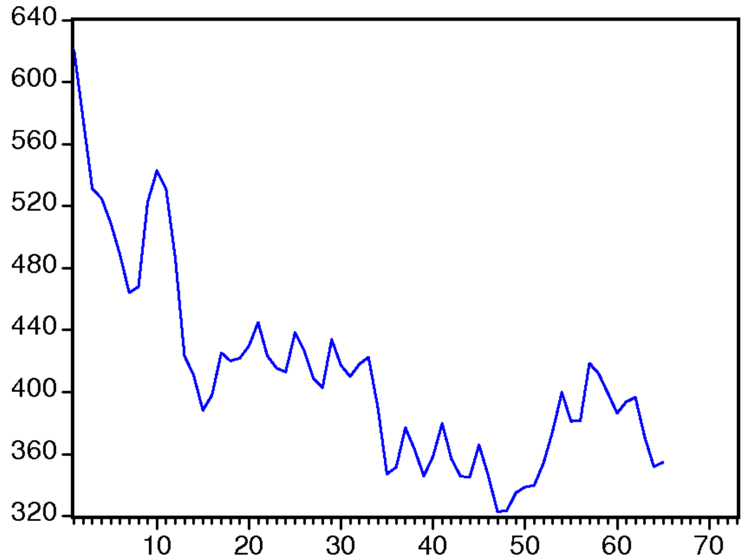


Figure 33: Quarterly average fares for Delta Airlines ($ against time).

which is a little higher than for the first models for the airline, but much lower than with a dummy variable in the regression.

- **Northwest:** Looking at the graph of the quarterly fares for Northwest Airlines (Figure 35), a clear trend appears as the fares seem to continuously decrease over the 15 last years. This trend may also explain the recent merger between Northwest and Delta (completed in 2008), as Northwest airlines has tried to reduce its costs and expenses since 2001. Indeed, we can observe that the establishment of a statistical model for the fares of the airline is rather complicated without a coefficient of trend, for the filtered data (we get the model ARMA(3,1) or non-filtered data (we get the model ARMA(2,3)). Moreover these models show problems of stationarity and autocorrelation. When we include a trend in the model, we get a better model for these fares. The non-filtered data and the filtered data give us an ARMA(4,3) model. This model has a coefficient of determination R^2 equal to 0.95 (0.96 without the

Dependent Variable: FARE3
Method: Least Squares
Date: 01/10/10 Time: 11:19
Sample (adjusted): 2 65
Included observations: 64 after adjustments
Convergence achieved after 8 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	462.6359	26.48296	17.46919	0.0000
@TREND	-1.826865	0.656994	-2.780643	0.0072
AR(1)	0.705859	0.085449	8.260588	0.0000
MA(1)	0.493513	0.122755	4.020321	0.0002

R-squared	0.907808	Mean dependent var	408.5544
Adjusted R-squared	0.903198	S.D. dependent var	58.40192
S.E. of regression	18.17056	Akaike info criterion	8.697943
Sum squared resid	19810.14	Schwarz criterion	8.832874
Log likelihood	-274.3342	F-statistic	196.9387
Durbin-Watson stat	1.983053	Prob(F-statistic)	0.000000

Inverted AR Roots	.71
Inverted MA Roots	-.49

Figure 34: Delta Airlines - regression on quarterly average fares.

filter). There is no stationarity problem according to the unit root test, and the residuals are not correlated. Looking at the Schwarz criterion, we can observe that this statistic is lower for the non-filtered data, so we decide to keep this model (Figure 36). If we include a dummy variable in the regression, we can observe that the autocorrelation problems remain in the regression (we get the model ARMA(0,2)).

- **United:** For United Airlines (real fares in Figure 37), we find with the same methods used before that the model ARMA(0,2) with a dummy variable (to take into account a drop of the fares at the period 35) seems to be the best for this carrier (Figure 38) for the filtered data. The model is stable, the coefficients are all individually significant, and the correlograms of the regression do not show any spikes in the autocorrelation coefficients, so this model seems to be quite a good one. The coefficient of determination of the regression is equal

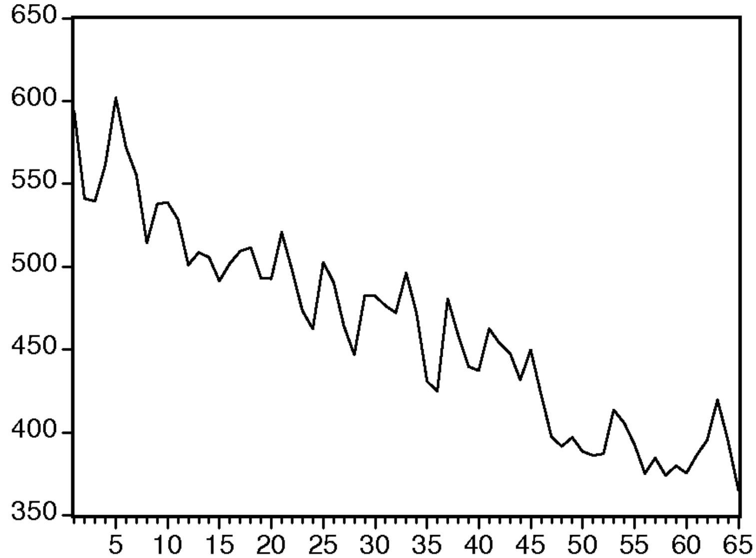


Figure 35: Quarterly average fares for Northwest Airlines ($ against time).

to 0.85, which is quite high. The dummy variable is significant in the model ARMA(0,2), and reveals a drop of about $78 in the fares at the time of the U.S. terrorist attacks. For the non-filtered data, EViews6 gives the model ARMA(1,2) with our dummy variable as the best model. This model is also stationary, and seems to be a good one. Nevertheless, the Schwarz criterion is higher for this regression than for the filtered data. Thus we can keep the filtered model for our analysis. If we do not include a dummy variable in the regression, the best model (according to the Schwarz Information Criterion) appears to be ARMA(3,2) for the filtered data, and ARMA(3,0) without the filter. These models appears to be good models too, but the use of a dummy variable decreases the SIC criterion. The use of a trend also could improve the model, but it leads to problems of autocorrelation with a quite low Durbin-Watson statistic.

- **US Airways:** For US Airways (real fares in figure 39), the same methods lead to the model ARMA(0,2) with a trend for the filtered data. This model seems to be stable, with nice t-

Dependent Variable: FARE4
Method: Least Squares
Date: 01/08/10 Time: 16:38
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 76 iterations
Backcast: 2 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	556.4219	8.218436	67.70410	0.0000
@TREND	-2.865983	0.210066	-13.64325	0.0000
AR(1)	-0.494535	0.144094	-3.432030	0.0012
AR(2)	-0.395445	0.131545	-3.006146	0.0041
AR(3)	-0.393415	0.124337	-3.164100	0.0026
AR(4)	0.469012	0.133880	3.503215	0.0010
MA(1)	1.427952	0.105101	13.58651	0.0000
MA(2)	1.273337	0.148822	8.556115	0.0000
MA(3)	0.813511	0.081167	10.02267	0.0000

R-squared	0.961141	Mean dependent var	459.6851
Adjusted R-squared	0.955163	S.D. dependent var	55.37599
S.E. of regression	11.72578	Akaike info criterion	7.896909
Sum squared resid	7149.684	Schwarz criterion	8.208349
Log likelihood	-231.8557	F-statistic	160.7710
Durbin-Watson stat	1.887232	Prob(F-statistic)	0.000000

Inverted AR Roots	.53	-.02-.95i	-.02+.95i	-.99
Inverted MA Roots	-.23+.88i	-.23-.88i	-.98	

Figure 36: Northwest Airlines - regression on quarterly average fares.

statistics for all coefficients, and there are no spikes in the correlogram graphs. The coefficient of determination here is equal to 0.88 (see Figure 40). Without the use of this trend variable, the best ARMA model we get is ARMA(3,2), which is also a good, but more complicated model. Moreover the use of a dummy variable is less relevant here than for United Airlines (according to the Schwarz Criterion). The use of the whole database leads to the model ARMA(3,3), and is not improved by the use of a dummy variable or a trend.

- **Southwest:** For the low-cost airline (real fares in Figure 41), the best ARMA model given by EViews6 according to the Schwarz criterion is an ARMA(0,3) with dummy variable and the trend variable for the filtered and non-filtered data. As for United and US Airways, the

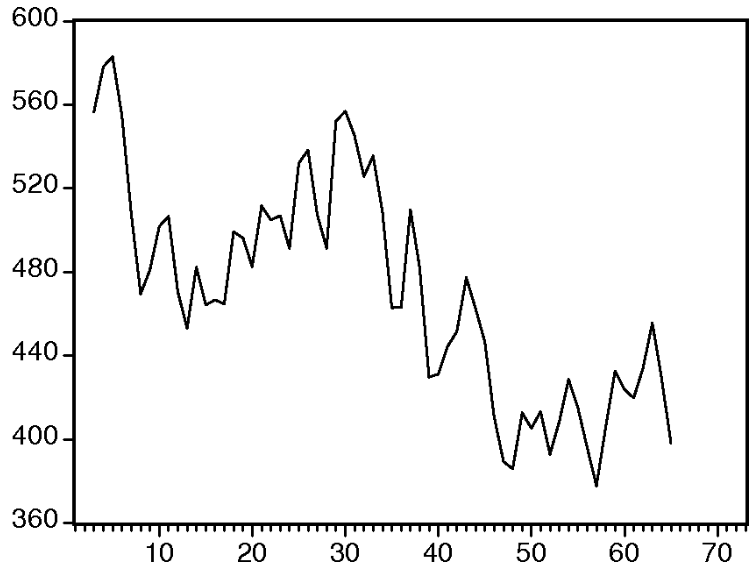


Figure 37: Quarterly average fares for United Airlines ($ against time).

Schwarz criterion is higher for the non-filtered data, thus we will keep the regression output for the filtered data (Figure 142). The coefficient of determination is equal to 0.84, the model is stable, all the coefficients matter in the regression, and there is no spike in the correlogram graphs (despite a quite low Durbin-Watson statistic, equal to 1.63).

The use of simple ARMA models leads here to the AR(1) model as the best one for the filtered or non-filtered data. These models are transformed into ARMA(0,1) if we include a dummy variable or a trend coefficient in the regressions, but these models have a lower SIC criterion than the ARMA(0,3) model with a dummy variable and a trend. We can notice that the MA roots for this model are very close to 1, which could lead to problems of non-invertibility for the fares of Southwest.

Dependent Variable: FARE5
Method: Least Squares
Date: 01/04/10 Time: 16:49
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 13 iterations
Backcast: 1 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	506.1110	9.798012	51.65446	0.0000
DUMMY	-77.86129	13.35465	-5.830275	0.0000
MA(1)	1.064661	0.092880	11.46275	0.0000
MA(2)	0.673172	0.092301	7.293217	0.0000

R-squared	0.847939	Mean dependent var	470.2444
Adjusted R-squared	0.840207	S.D. dependent var	52.41618
S.E. of regression	20.95287	Akaike info criterion	8.983816
Sum squared resid	25902.35	Schwarz criterion	9.119888
Log likelihood	-278.9902	F-statistic	109.6677
Durbin-Watson stat	1.981516	Prob(F-statistic)	0.000000
Inverted MA Roots	-.53+.62i	-.53-.62i	

Figure 38: United Airlines - regression on quarterly average fares.

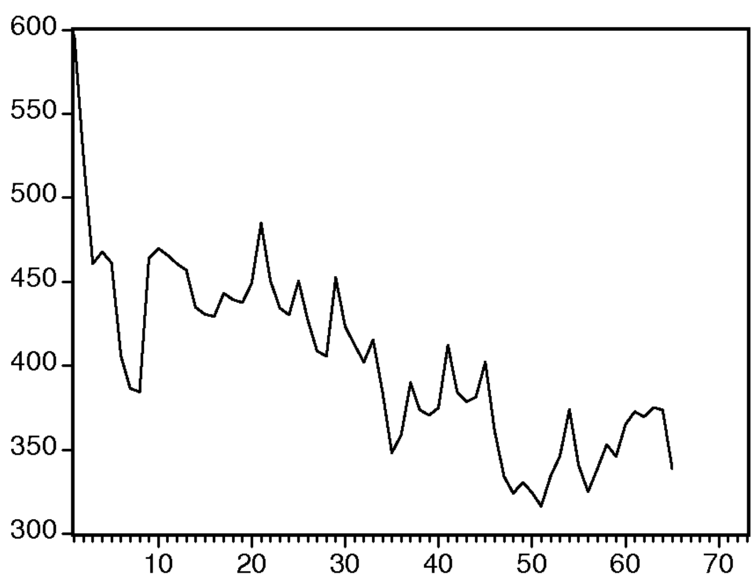


Figure 39: Quarterly average fares for US Airways ($ against time).

Dependent Variable: FARE6
Method: Least Squares
Date: 01/18/10 Time: 18:17
Sample: 1 65
Included observations: 65
Convergence achieved after 9 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	465.1703	12.42576	37.43596	0.0000
@TREND	-2.064267	0.327728	-6.298722	0.0000
MA(1)	1.007156	0.073972	13.61534	0.0000
MA(2)	0.611885	0.053711	11.39222	0.0000

R-squared	0.884305	Mean dependent var	402.6931
Adjusted R-squared	0.878615	S.D. dependent var	53.94896
S.E. of regression	18.79599	Akaike info criterion	8.764727
Sum squared resid	21550.64	Schwarz criterion	8.898536
Log likelihood	-280.8536	F-statistic	155.4164
Durbin-Watson stat	2.147299	Prob(F-statistic)	0.000000

Inverted MA Roots -.50+.60i -.50-.60i

Figure 40: US Airways - regression on quarterly average fares.

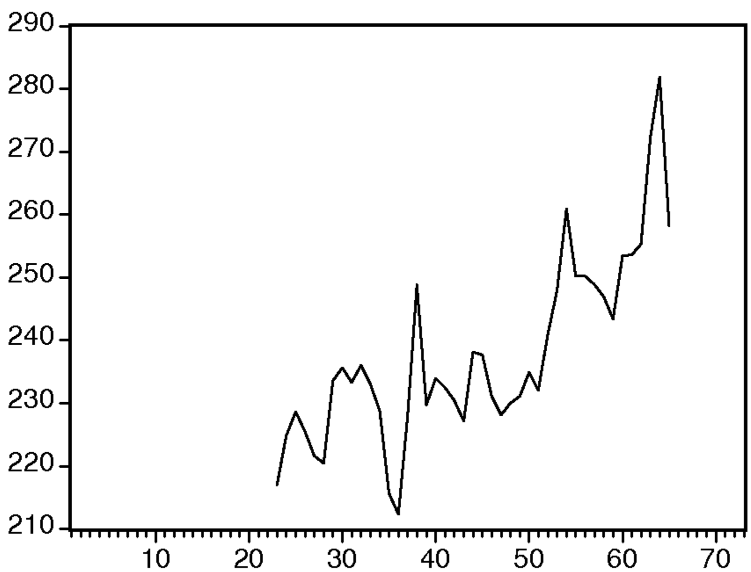


Figure 41: Quarterly average fares for Southwest Airlines ($ against time).

Dependent Variable: FARE7
Method: Least Squares
Date: 01/18/10 Time: 18:28
Sample (adjusted): 23 65
Included observations: 43 after adjustments
Convergence achieved after 21 iterations
Backcast: 20 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	191.2149	2.219953	86.13462	0.0000
DUMMY	-15.54111	2.749888	-5.651542	0.0000
@TREND	1.321151	0.084972	15.54803	0.0000
MA(1)	0.507582	0.040512	12.52921	0.0000
MA(2)	-0.485910	0.032115	-15.13021	0.0000
MA(3)	-0.900402	0.054744	-16.44746	0.0000

R-squared	0.844654	Mean dependent var	237.7842
Adjusted R-squared	0.823661	S.D. dependent var	14.63999
S.E. of regression	6.147738	Akaike info criterion	6.598833
Sum squared resid	1398.403	Schwarz criterion	6.844582
Log likelihood	-135.8749	F-statistic	40.23547
Durbin-Watson stat	1.629990	Prob(F-statistic)	0.000000

Inverted MA Roots	.96	-.74-.63i	-.74+.63i

Figure 42: Southwest Airlines - regression on quarterly average fares.

4.1.5 Summing up

Table 2 below summarizes the "best" models we get from our statistical regressions in EViews6. We can observe that there is no need for any carrier to use the differencing process to get the best model. The ARMA(0,2) model seems to be adopted by 4 carriers: American Airlines, Continental Airlines, United Airlines and US Airways. Among these carriers, the average fares of American Airlines seem to decrease since 1993, by about \$16.5 every 10 years, which is not so much for air transport fares. In addition to this trend, and as for the 3 other carriers, the fares of American Airlines seem to have been influenced by the terrorist attacks of September 11, 2001. Indeed, EViews6 shows a drop in these fares, at the time of these attacks, of about \$88 for American, \$52 for Continental, \$78 for United and \$65 for US Airlines. We can assume that for the other carriers, they may have been less affected by these attacks than American Airlines, or may have had a different strategy than its rival.

We can explain the use of ARMA(0,q) models by the use of a trend variable in many models, which reduces the importance of an autoregressive component in these models. These trends for the fares to decrease or increase over a period of 15 years may in part be explained by macroeconomic data. We will see that below. Indeed, we can observe that there is a trend for the fares of Northwest to decrease (by about \$29 every ten years), which may explain the situation of crisis recently undergone by the carrier before it was merged with Delta Airlines. On the contrary, here we can notice the obvious difference between the legacy carriers and Southwest, which is the only airline to progressively increase its fare (by about \$14 every ten years), even if the low-cost airline also suffered a sudden drop of its fares in 2001 (but of only \$16, in comparison with \$88 for American Airlines, for example).

All these models are stable, with no problems of autocorrelation in the residuals, high significance of each coefficient of the regressions and also high significance of all coefficients (given with the F-statistic of the regressions which follows a t-statistic). The Jarque-Bera statistic is also quite

Carriers	**AA**	**CO**	**DL**	**NW**	**UA**	**US**	**WN**
use of filter on data	no	no	yes	no	yes	yes	yes
AR degree	0	0	1	4	0	0	0
MA degree	2	2	1	3	2	2	3
differencing	no	no	no	no	no	no	no
ARIMA model	(0,0,2)	(0,0,2)	(1,0,1)	(4,0,3)	(0,0,2)	(0,0,2)	(0,0,3)
trend variable	yes	no	yes	yes	no	yes	yes
value of trend coefficient	-1.65		-1.83	-2.87		-2.06	1.32
dummy variable at period 35	yes	yes	no	no	yes	no	yes
value of dummy coefficient	-88	-52			-78		-16
R^2	0.93	0.73	0.91	0.96	0.85	0.88	0.84
Schwarz criterion	9.20	9.22	8.83	8.21	9.12	8.90	6.84
Durbin-Watson statistic	2.22	2.20	1.98	1.89	1.98	2.15	1.63
Jarque-Bera statistic	4.99	2.13	0.71	2.51	4.49	0.20	0.40
F-statistic	214	56	197	161	110	155	40

Table 2: The carriers are reported with their IATA airline codes.

low for all the models, meaning that the residuals of all the models approximately follow a normal distribution.

4.1.6 Forecasting

Once we get these 7 ARIMA models for our carriers, we may wonder if we are able to infer the future average fares for an airline? A first idea in order to ensure the reliability of our model could be to "forecast the past" (in-sample). Indeed, we could try to apply our "best" model for each airline at a past period of time (for example from the first quarter in 2007), as if it was our last period of data, and compare this model for 2 years (until 2009) to the data we have for these 2 years. We get the following graphs (see Figure 43): we can observe 4 curves on these graphs, between quarter 58 and quarter 65: the real data, plotted for all the periods of observation (65 quarters), the best model for these fares (dashed curve), and around this forecast, the possible distribution of these fares at the 95% level of confidence around the value of the modelled fares. This distribution is computed as follows: at a given period t, for our model, and at a given level of confidence α, we have

$$\widehat{\text{Fare}}(t) - (\text{t} - \text{stat}_\alpha(n) * \varepsilon(t)) \leq \text{Fare}(t) \leq \widehat{\text{Fare}}(t) + (\text{t} - \text{stat}_\alpha(n) * \varepsilon(t)) \tag{25}$$

, where $\text{t} - \text{stat}_\alpha(n)$ is the t-statistic at the α level of confidence for 2-tailed testing. Here, since $n = 65$, $\text{t} - \text{stat}_\alpha(n) \approx 2$. $\widehat{\text{Fare}}(t)$ is the estimated fare at time t according to the best model for our regression.

We can observe that the trend for all the fares is quite well represented by our forecast. The real data are caught within the bounds of the standard errors for this model. However, in most cases, the model is not able to catch the short spikes in the real data and the forecast softens quite quickly.

We can also use this method to forecast the fares after the 65 periods of observations (out-of-sample). If we use the best model for each carrier and apply this model during 8 periods after the period 65, we get the following graphs (Figure 44). As we saw in 43, the forecast quickly softens after 3 or 4 quarters of modelled fares. This is a consequence of the stationarity of our models.

We can observe that the model foresees decreasing trends of the fares for American Airlines, Delta Air Lines, Northwest Airlines and US Airways. For Continental Airlines and United Airlines, the fares seems to converge to a limit (about $420 for both carriers). And as we saw in the regressions for Southwest airlines, the model for the low-cost airline predicts an important increasing trend for the average fares.

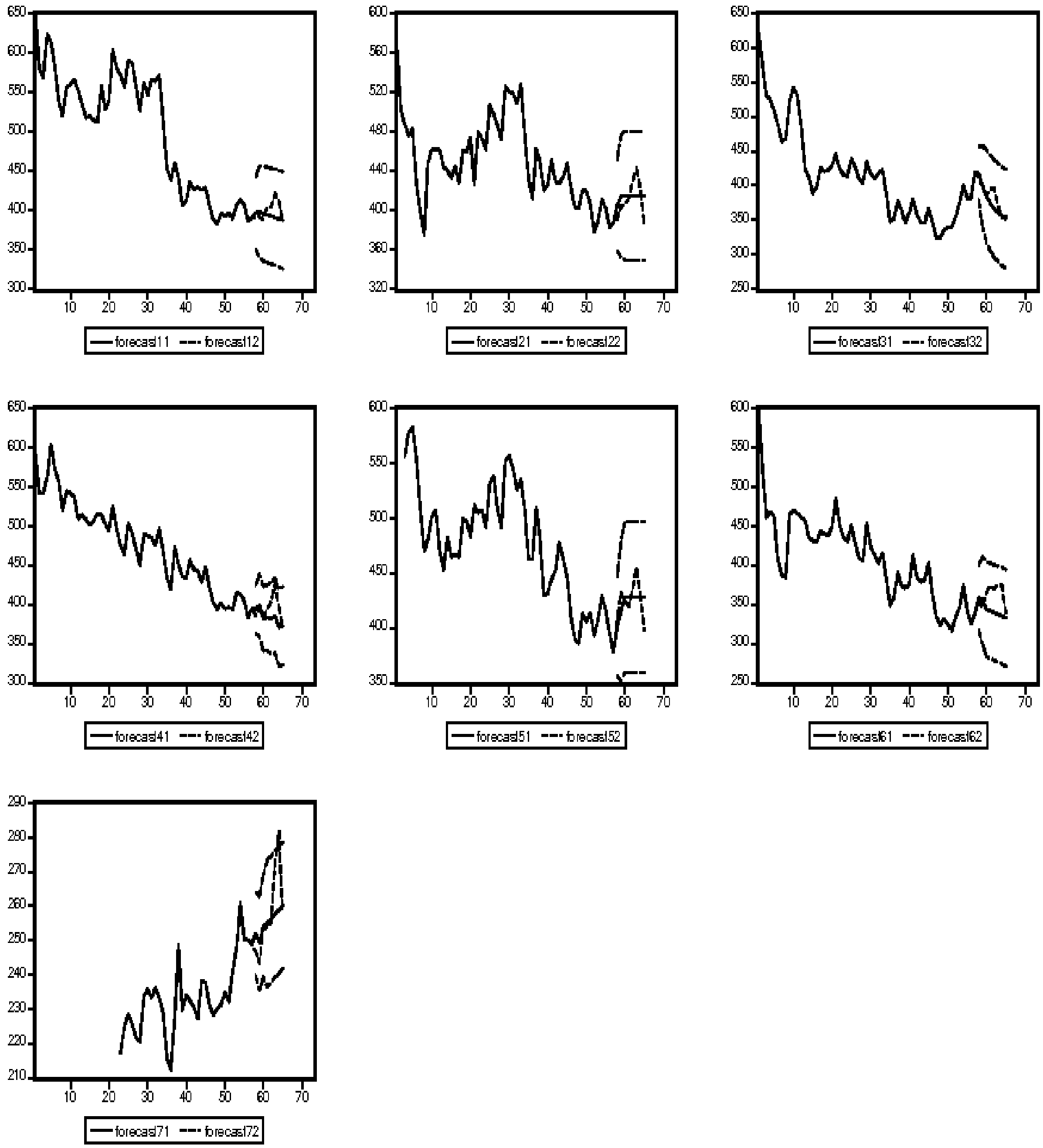


Figure 43: In-sample dynamic forecasts of average fares from quarter 58 to quarter 65. Forecast 1 refers to American, forecast 2 refers to Continental, 3 refers to Delta, 4 refers to Northwest, 5 refers to United, 6 refers to US Airways and 7 refers to Southwest.

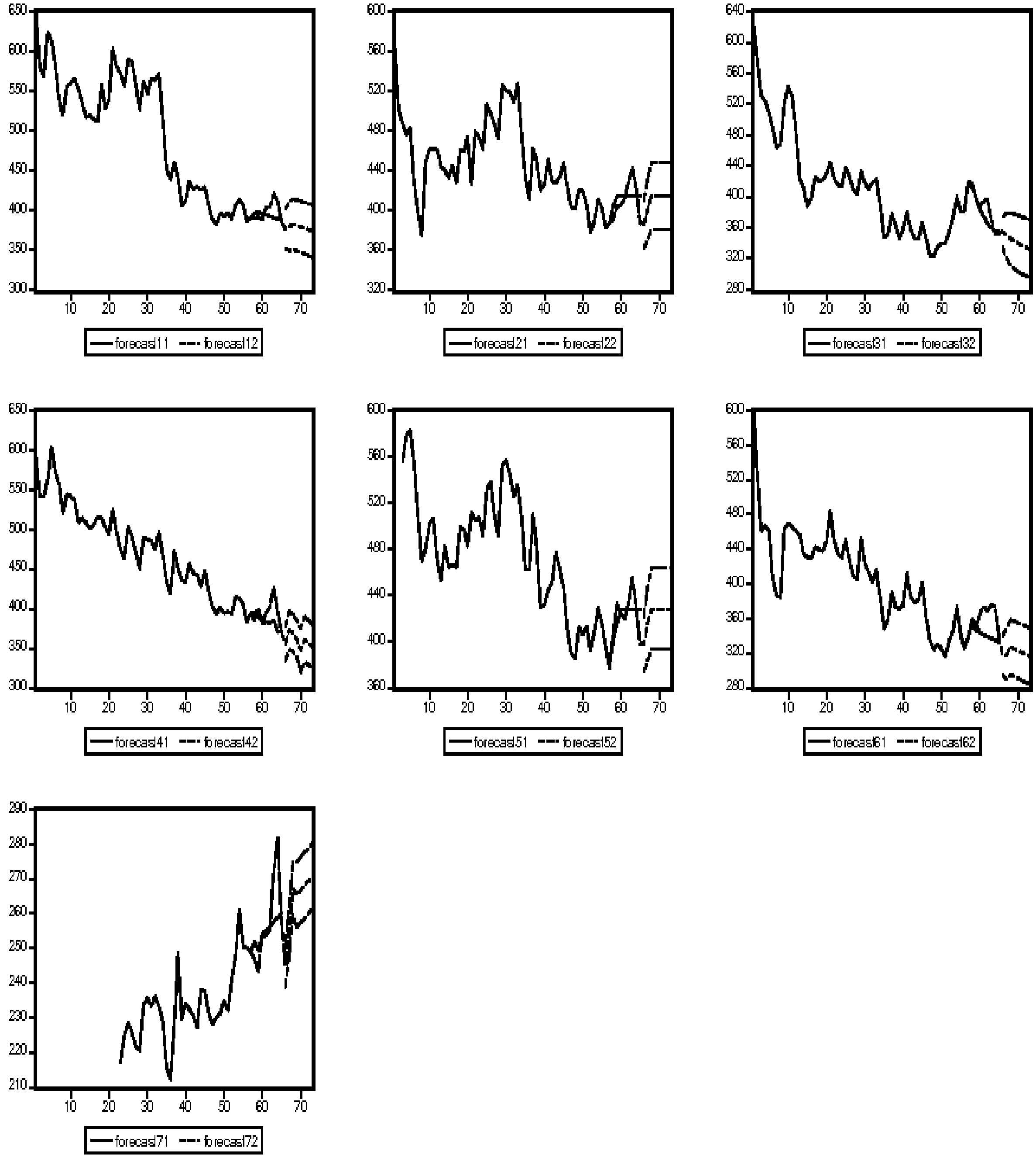


Figure 44: Out-of-sample dynamic forecasts of average fares from quarter 65 to quarter 73. Forecast 1 refers to American, forecast 2 refers to Continental, 3 refers to Delta, 4 refers to Northwest, 5 refers to United, 6 refers to US Airways and 7 refers to Southwest.

4.2 Passengers

We can now try to find models for the number of passengers (for domestic round trips in the U.S.) of these same 7 airlines. We first look at the graphs of this statistic (Figure 45). We notice that

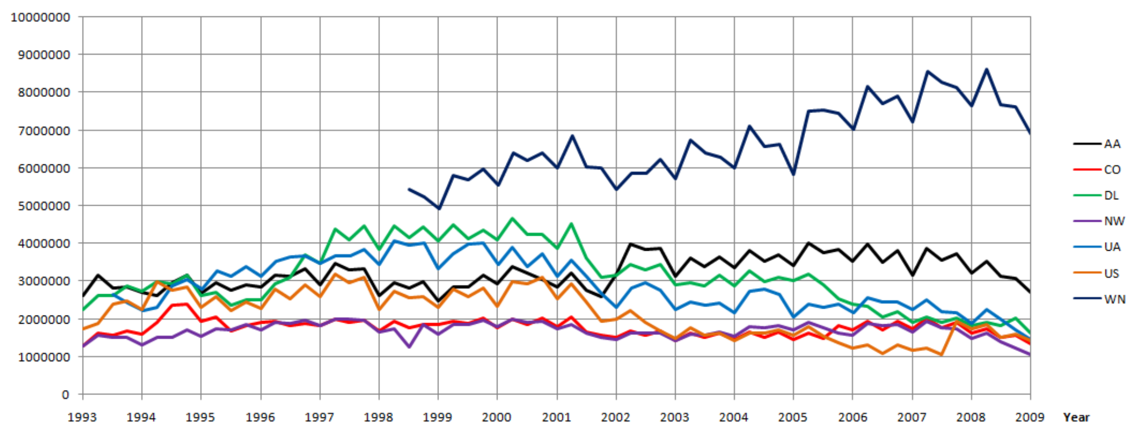


Figure 45: Number of passengers carried per quarter (on round trips) from 1993 (quarter 1) to 2009 (quarter 65).

the number of passengers for these airlines seems to include a quarterly component of seasonality. Indeed, we saw in the first part of the report that the number of passengers flying during the second quarter of the year is significantly higher than during the first or the fourth quarter of the year. A good model will need to take that into account. In order to catch this seasonality of fares for several airlines, we can add a seasonal autoregressive component in the regressions. Consider for example an ARMA(2,2) model :

$$\text{Fare}(t) = \alpha_0 + \alpha_1 \times \text{Fare}(t-1) + \alpha_2 \times \text{Fare}(t-2) + \varepsilon(t) + \beta_1 \times \varepsilon(t-1) + \beta_2 \times \varepsilon(t-2). \qquad (26)$$

This model can also be written, using the lag operator L,

$$A(L) \times \text{Fare}(t) = \alpha_0 + B(L) \times \varepsilon(t). \qquad (27)$$

Then, the insertion of a seasonal autoregressive lag of order 4 (SAR(4)) can be included:

$$A(L) * (1 - \gamma_0 \times L^4) \times \text{Fare}(t) = \alpha_0 + B(L) \times \varepsilon(t). \tag{28}$$

The parameter γ_0 is associated with the seasonal component of the process. This method can also be used for the MA part of the model, with the use of a seasonal moving average component of order 4 (SMA(4)) in the regression, using the product of the polynomial $B(L)$ with the lag polynomial $1 + \delta_0 \times L^4$. The use of a lag of order 4 for the seasonal component of the models seems to be here the most natural solution for our models, as the period of seasonality for the number of passengers appear to be one year (4 quarters).

If we run a first series of simple ARMA regressions, allowing our program in EViews6 to use a seasonal AR component of order 4 if it can reduce the Schwarz Information Criterion and, as a consequence, maybe improve our model, we get the following results:

- **American:** For American Airlines, even if the number of passengers appears to be seasonal (Figure 46), EViews6 chooses an ARMA(2,1) model without a seasonal component as the best model it can find (Figure 47). Nevertheless, we can observe from the regression output that this model is non-stationary and non-invertible, with an AR root higher than 1 in absolute value and a MA root equal to 1 in absolute value. If we include our seasonal component SAR(4) (Figure 48), even if the Schwarz Information Criterion is higher than for the first regression, this new coefficient AR(4) leads to a stationary and invertible model. Nevertheless, not all the coefficients in the regression appear to be individually significant. The Durbin Watson statistic is also quite low (1.92), and we can observe from the correlograms that there is some autocorrelation between the residuals of the regression. The R^2 statistic is also quite low (0.59), so we can infer that this model can be improved. We saw that it is also possible to use a seasonal MA component (also of order 4) in our models. If we include

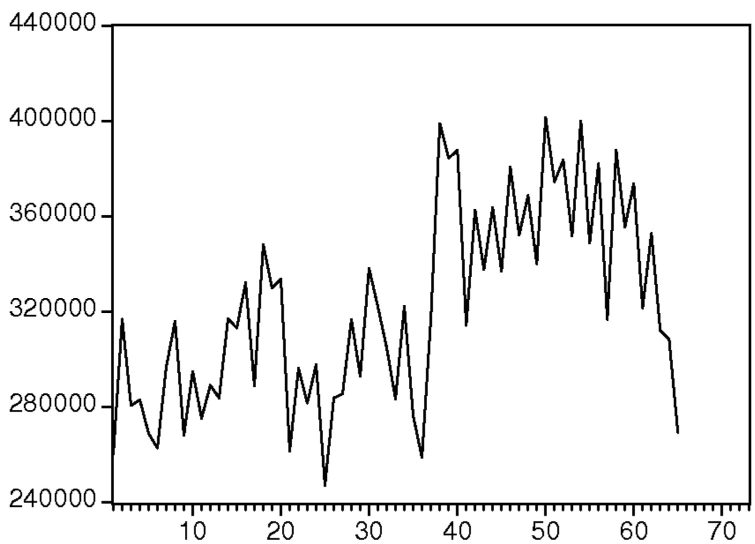


Figure 46: Quarterly number of passengers carried by American Airlines on round trips.

a SMA(4) coefficient in the regression, we obtain a non-stationary and maybe non-invertible model (Figure 49). If we impose the use of the seasonal component AR(4), and run the program in order to keep the model with the lowest SIC, we obtain an ARMA(0,2) model (with the seasonal component SAR(4)) (Figure 50). We can observe that all the coefficients are individually and jointly significant, and that the model is invertible and stationary. The Durbin Watson statistic is still quite low, but we can check on the correlograms that there is no autocorrelation in the residuals at the 95% level of confidence. We obtain an R^2 statistic equal to 0.62.

Remark: As for the regressions on the quarterly average fares of the airline, these regressions on the number of passengers use the filtered data of our database. However, if we use the non-filtered data, the result appears to be an ARMA(0,2) with the seasonal component SAR(4) too (Figure 51). The R^2 statistic and the SIC are nonetheless not as good as for the regression with the filtered data.

Dependent Variable: PAX1
Method: Least Squares
Date: 01/21/10 Time: 09:49
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 28 iterations
Backcast: 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	322180.7	16902.76	19.06083	0.0000
AR(1)	-0.192253	0.078517	-2.448555	0.0173
AR(2)	0.838684	0.077086	10.87980	0.0000
MA(1)	0.997385	0.045562	21.89049	0.0000

R-squared	0.669609	Mean dependent var	322969.0
Adjusted R-squared	0.652809	S.D. dependent var	41344.73
S.E. of regression	24361.52	Akaike info criterion	23.10078
Sum squared resid	3.50E+10	Schwarz criterion	23.23686
Log likelihood	-723.6747	F-statistic	39.85870
Durbin-Watson stat	1.870131	Prob(F-statistic)	0.000000

Inverted AR Roots .82 -1.02
Estimated AR process is nonstationary
Inverted MA Roots -1.00

Figure 47: American Airlines - regression on the quarterly number of passengers (filtered data).

- **Continental:** Using the same methodology, we also find out that if we do not impose the use of the coefficient SAR(4) in our regression, the model we get (an ARMA(3,4) with our dummy variable and without the coefficient SAR(4)) is non-stationary and non-invertible (for the filtered and non-filtered data). However, we notice that, even if we include an SAR(4) variable or SMA(4) variable, we still have problems of stationarity for both databases.

  Looking at the graph of the number of passengers for the carrier (Figure 52), we can observe that there seems to be a gap in the number of passenger in the period 35 (presumably related to the September 11 attacks). Thus if we try to impose the use of our dummy variable in the model for Continental Airlines, the model we get with the filtered data appears to be very complex (an ARMA(2,4) with a dummy variable and a seasonal component SAR(4)). However, if we use the complete database, we get an ARMA(0,2) model with our dummy

Dependent Variable: PAX1
Method: Least Squares
Date: 01/21/10 Time: 10:12
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 228 iterations
Backcast: 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	333135.7	28234.90	11.79872	0.0000
AR(1)	0.573187	0.406710	1.409327	0.1645
AR(2)	-0.071156	0.277897	-0.256051	0.7989
SAR(4)	0.714231	0.105706	6.756761	0.0000
MA(1)	0.137400	0.428566	0.320604	0.7497

R-squared	0.592411	Mean dependent var	326291.7
Adjusted R-squared	0.562220	S.D. dependent var	40566.20
S.E. of regression	26840.61	Akaike info criterion	23.31416
Sum squared resid	3.89E+10	Schwarz criterion	23.49022
Log likelihood	-682.7677	F-statistic	19.62164
Durbin-Watson stat	1.917247	Prob(F-statistic)	0.000000

Inverted AR Roots	.92	.39	.18	.00-.92i
	.00+.92i	-.92		
Inverted MA Roots	-.14			

Figure 48: American Airlines - regression on the quarterly number of passengers (filtered data).

variable and the seasonal component SAR(4) (see Figure 53), which appears to be quite a good model. This model is stationary, so we can infer that the use of this dummy variable is a good solution for Continental Airlines. Indeed, we can observe that all the coefficients are significant, with a coefficient for the dummy variable equal to $-4{,}8710$, which means that according to this model the number of passengers dropped by about 50,000 in the period of the U.S. terrorist attacks. The Durbin Watson statistic is quite low, but there is no apparent spike in the correlograms of the regressions, so the residuals appear not to be autocorrelated. The R^2 statistic is very high (0.81) compared to the statistic we get with the model for American Airlines.

- **Delta Air Lines:** For Delta Air Lines, the use of seasonal components in the regression

Dependent Variable: PAX1
Method: Least Squares
Date: 01/21/10 Time: 10:17
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 21 iterations
Backcast: 2 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	326285.2	180699.2	1.805682	0.0766
AR(1)	0.281747	0.167523	1.681842	0.0985
AR(2)	0.406260	0.167297	2.428377	0.0186
SAR(4)	1.027199	0.025439	40.37888	0.0000
MA(1)	0.570184	0.136224	4.185633	0.0001
SMA(4)	-0.916045	0.069748	-13.13361	0.0000

R-squared	0.722963	Mean dependent var	326291.7
Adjusted R-squared	0.696828	S.D. dependent var	40566.20
S.E. of regression	22336.20	Akaike info criterion	22.96195
Sum squared resid	2.64E+10	Schwarz criterion	23.17322
Log likelihood	-671.3775	F-statistic	27.66205
Durbin-Watson stat	1.656182	Prob(F-statistic)	0.000000

Inverted AR Roots	1.01	.79	.00+1.01i	-.00-1.01i
	-.51	-1.01		
	Estimated AR process is nonstationary			
Inverted MA Roots	.98			

Figure 49: American Airlines - regression on the quarterly number of passengers (filtered data).

appears to make the model unstable, and EViews6 does not manage to find a stationary ARMA model with a seasonal component. If we look at the graph of the number of passengers for the airline (Figure 54), we can observe that, as for Continental Airlines, there was an important drop in the number of passengers carried by Delta in period 35. Thus, as we could expect according to this graph, if we do not use the seasonal coefficients in our regression, EViews6 gives the model AR(3) with a dummy variable as the best model according to the Schwarz Information Criterion, for the filtered data (Figure 55). For the non-filtered data, EViews6 gives a ARMA(2,1) with a dummy variable and a trend variable (Figure 56). Both models are stationary, but the model for non-filtered data is not invertible, so we may keep the regression output for the filtered data. All coefficients are individually significant,

Dependent Variable: PAX1
Method: Least Squares
Date: 01/21/10 Time: 10:41
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 13 iterations
Backcast: 3 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	326588.4	15818.06	20.64655	0.0000
AR(4)	0.555991	0.113867	4.882813	0.0000
MA(1)	0.499125	0.102449	4.871933	0.0000
MA(2)	0.606607	0.098701	6.145918	0.0000

R-squared	0.623211	Mean dependent var	324314.0
Adjusted R-squared	0.603380	S.D. dependent var	41332.54
S.E. of regression	26030.31	Akaike info criterion	23.23524
Sum squared resid	3.86E+10	Schwarz criterion	23.37365
Log likelihood	-704.6747	F-statistic	31.42613
Durbin-Watson stat	1.766312	Prob(F-statistic)	0.000000

Inverted AR Roots	.86	
Inverted MA Roots	-.25-.74i	-.25+.74i

Figure 50: American Airlines - regression on the quarterly number of passengers (filtered data).

with a high coefficient of determination for both regressions. The Durbin-Watson statistic is quite low as for the previous airlines, but there are no spikes to indicate possible problems of autocorrelation in the correlograms. For the model obtained with the filtered data, we can notice that the drop in the number of passengers after the period of the U.S. terrorist attacks is about 135,000 passengers, which is roughly much 30% of the total number of passengers carried by Delta in period 34!

- **Northwest Airlines:** It also appears difficult to avoid the problems of stationarity for the number of passengers carried quarterly by Northwest Airlines (Figure 57). The best model we can obtain with all the possible parameters in our program is an ARMA(2,1) with a dummy variable and no correction for possible seasonality of the fares. However, this model appears to be non-stationary. If we do not filter the data, EViews6 proposes the same model

Dependent Variable: PAX1
Method: Least Squares
Date: 01/21/10 Time: 10:48
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 12 iterations
Backcast: 3 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	380058.1	14059.06	27.03296	0.0000
AR(4)	0.428841	0.125038	3.429687	0.0011
MA(1)	0.519392	0.106697	4.867896	0.0000
MA(2)	0.546597	0.108486	5.038411	0.0000

R-squared	0.453922	Mean dependent var	384009.7
Adjusted R-squared	0.425181	S.D. dependent var	39874.47
S.E. of regression	30231.56	Akaike info criterion	23.53449
Sum squared resid	5.21E+10	Schwarz criterion	23.67290
Log likelihood	-713.8018	F-statistic	15.79355
Durbin-Watson stat	1.842834	Prob(F-statistic)	0.000000

Inverted AR Roots	.81	.00-.81i	.00+.81i	-.81
Inverted MA Roots	-.26+.69i	-.26-.69i		

Figure 51: American Airlines - regression on the quarterly number of passengers (non-filtered data).

as the best one (according to the SIC), but with a seasonal coefficient SAR(4) (Figure 58). This model is stationary, but non-invertible. All coefficients are individually significant, the R^2 statistic is quite high (0.80), but there are still some problems of autocorrelation in the residuals (we can observe a spike in the correlograms at the 6th lag).

- **United Airlines:** For United Airlines (quarterly number of passengers plotted in Figure 59), EViews6 gives the ARMA(0,2) with a dummy variable, a trend variable and the seasonal variable SAR(4) (here equal to the variable AR(4) as there is no other AR lag in the regression) (Figure 60). For the non-filtered data, EViews6 adds an AR lag of order 1 to the regression to find the best model for the number of passengers of the airline (Figure 61). Both models seem to be quite good: all coefficients are individually (and jointly) significant, stationary, invertible, with no problems of autocorrelation. The R^2 statistic is quite high (0.93 and 0.94),

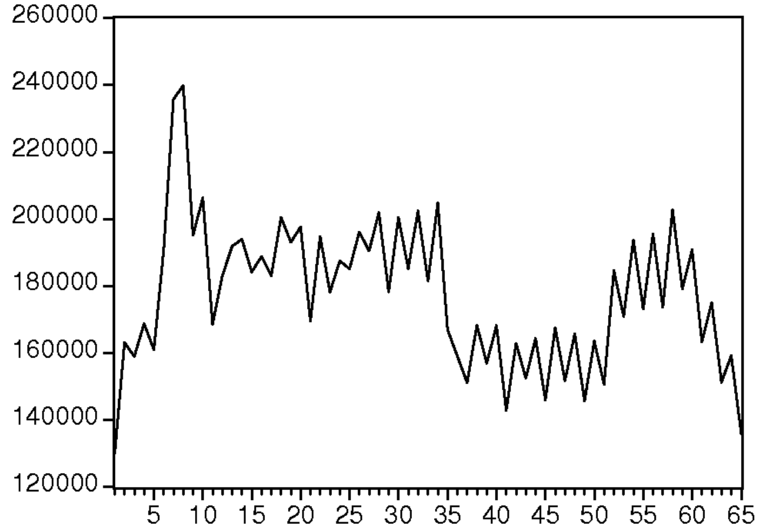


Figure 52: Number of passengers carried by Continental Airlines per quarter.

the Schwarz criterion is a little lower for the regression on the filtered data (the different number of explanatory variables in the regression can explain that the SIC is higher for the regression on the whole database). These regressions show an important trend for the number of passengers of the airline to decrease over time, by more or less 4000 passengers per quarter. United Airlines also suffered an important drop in its number of passengers in 2001, of about 59,000 passengers for the filtered database and 70,000 passengers without the lower bound on the data (lower bound for all flights with less than 90 passengers, which explains the difference between the coefficients of the dummy variables).

- **US Airways:** For US Airways (quarterly number of passengers plotted in Figure 62), the best model we get is an ARMA(2,1) with a dummy variable, for the filtered data (Figure 63).

  We can observe that this regression also has problems with stationarity. In order to make our model stationary, we include in the model the seasonal coefficient SAR(4). According to the SIC, the best model we can then get is an ARMA(1,2) with a dummy variable, the

Dependent Variable: PAX2
Method: Least Squares
Date: 01/21/10 Time: 11:20
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 22 iterations
Backcast: 3 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	229040.4	6214.957	36.85309	0.0000
DUMMY	-48709.56	7348.126	-6.628841	0.0000
AR(4)	0.255780	0.109379	2.338471	0.0230
MA(1)	0.201643	0.025081	8.039579	0.0000
MA(2)	0.953083	0.026899	35.43151	0.0000

R-squared	0.807005	Mean dependent var	204695.8
Adjusted R-squared	0.793220	S.D. dependent var	28793.51
S.E. of regression	13093.30	Akaike info criterion	21.87600
Sum squared resid	9.60E+09	Schwarz criterion	22.04902
Log likelihood	-662.2181	F-statistic	58.54074
Durbin-Watson stat	1.406419	Prob(F-statistic)	0.000000

Inverted AR Roots	.71	.00-.71i
Inverted MA Roots	-.10+.97i	-.10-.97i

Figure 53: Continental Airlines - regression on the quarterly number of passengers (non-filtered data).

seasonal coefficient SAR(4) and a seasonal moving average component SMA(4) (Figure 64). For the non-filtered data, the best model with a seasonal component is the same model with a trend variable (this coefficient shows a trend of about -1922 passengers carried by the airline per quarter). This model is still non-invertible, but it is stationary. All the coefficients are individually significant, and the R^2 statistic is equal to 0.94. The SIC criterion is again higher for the more complicated model (regression with the trend variable), so we will keep the results for the filtered data.

- **Southwest:** For Southwest, there are some problems with stationarity for many models, so it is very difficult to compare an important number of regressions with our program. Looking at the graph of the number of passengers for the low-cost airline, we can observe that there is a

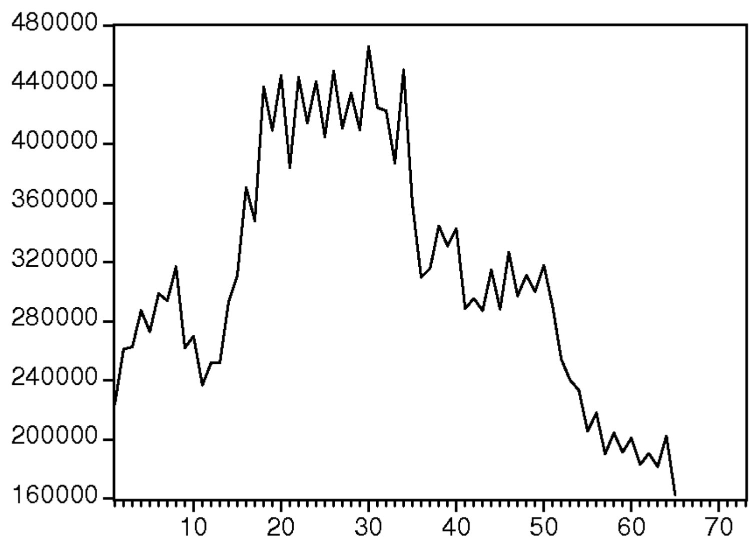


Figure 54: Number of passengers carried by Delta Air Lines per quarter.

Dependent Variable: PAX3
Method: Least Squares
Date: 01/21/10 Time: 12:27
Sample (adjusted): 4 65
Included observations: 62 after adjustments
Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	381400.0	41941.45	9.093630	0.0000
DUMMY	-134581.6	24922.08	-5.400095	0.0000
AR(1)	0.639650	0.116538	5.488761	0.0000
AR(2)	0.786474	0.101400	7.756157	0.0000
AR(3)	-0.503618	0.111887	-4.501142	0.0000

R-squared	0.927800	Mean dependent var	314346.4
Adjusted R-squared	0.922733	S.D. dependent var	84145.02
S.E. of regression	23389.76	Akaike info criterion	23.03519
Sum squared resid	3.12E+10	Schwarz criterion	23.20673
Log likelihood	-709.0909	F-statistic	183.1173
Durbin-Watson stat	1.793060	Prob(F-statistic)	0.000000

Inverted AR Roots	.89	.64	-.89

Figure 55: Delta Air Lines - regression on the quarterly number of passengers (filtered data).

Dependent Variable: PAX3
Method: Least Squares
Date: 01/21/10 Time: 12:32
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 89 iterations
Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	600737.5	65773.57	9.133418	0.0000
DUMMY	-120190.6	31818.37	-3.777397	0.0004
@TREND	-3631.146	1624.054	-2.235853	0.0293
AR(1)	-0.128164	0.046742	-2.741924	0.0081
AR(2)	0.837394	0.041679	20.09168	0.0000
MA(1)	1.156464	0.085352	13.54934	0.0000

R-squared	0.958505	Mean dependent var	401949.3
Adjusted R-squared	0.954865	S.D. dependent var	106588.0
S.E. of regression	22644.55	Akaike info criterion	22.98362
Sum squared resid	2.92E+10	Schwarz criterion	23.18773
Log likelihood	-717.9840	F-statistic	263.3331
Durbin-Watson stat	1.651653	Prob(F-statistic)	0.000000

Inverted AR Roots	.85	-.98
Inverted MA Roots	-1.16	
	Estimated MA process is noninvertible	

Figure 56: Delta Air Lines - regression on the quarterly number of passengers (non-filtered data).

significant trend for this statistic to increase over time (see Figure 65). Thus we can infer that we need to consider a model which takes into account this trend. Then we can first try to run a basic regression with a constant and a trend to observe the correlograms of the regression. We get the following regression output (Figure 66) and the associated correlograms (Figure 66). We can observe that the use of the trend is very important in this model, as we do not reject the coefficient at the 99% level of confidence. This trend is equal to +6533 passengers per quarter for the airline: this is clearly a sign of the strength of the low-cost airline in the U.S. domestic market, where almost all its "rivals" have faced a roughly stable or decreasing number of passengers over the past 15 years. The correlograms of the regression show a high autocorrelation with the fourth lag of the regression. We can assume that this is perhaps a

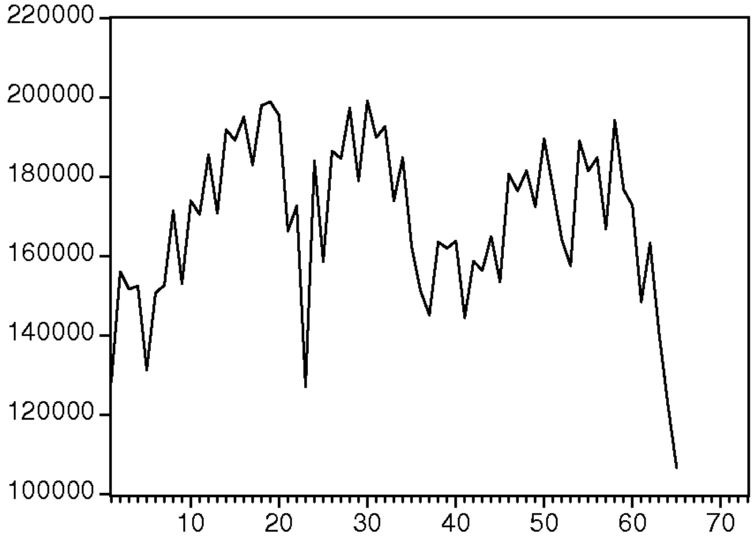


Figure 57: Number of passengers carried by Northwest Airlines per quarter.

sign of seasonality for the fares, so we can try to include a coefficient SAR(4) in the regression (Figure 68). We can see in the regression output that this model is better than the previous one (a higher R^2 statistic and a lower SIC). This model remains stationary, and is invertible. Nevertheless, there is an important spike which appeared in the correlograms (for the first lag of the residuals), so there are problems of autocorrelation in this regression, which does not include any ARMA term. In fact, we can observe that if we try to run our program with these coefficients of trend and seasonality, EViews6 is not able to find a better model than this one. The same problems also remain with the non-filtered data.

Dependent Variable: PAX4
Method: Least Squares
Date: 01/21/10 Time: 13:30
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 58 iterations
Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	254617.2	15436.95	16.49401	0.0000
DUMMY	-43080.92	17440.01	-2.470234	0.0168
AR(1)	-0.217220	0.100772	-2.155565	0.0357
AR(2)	0.751454	0.074553	10.07951	0.0000
SAR(4)	0.304807	0.135366	2.251722	0.0285
MA(1)	1.203116	0.139858	8.602432	0.0000

R-squared	0.804739	Mean dependent var	232577.6
Adjusted R-squared	0.786318	S.D. dependent var	35134.29
S.E. of regression	16241.08	Akaike info criterion	22.32462
Sum squared resid	1.40E+10	Schwarz criterion	22.53589
Log likelihood	-652.5763	F-statistic	43.68636
Durbin-Watson stat	1.799537	Prob(F-statistic)	0.000000

Inverted AR Roots	.77	.74
Inverted MA Roots	-1.20	
	Estimated MA process is noninvertible	

Figure 58: Northwest Airlines - regression on the quarterly number of passengers (non-filtered data).

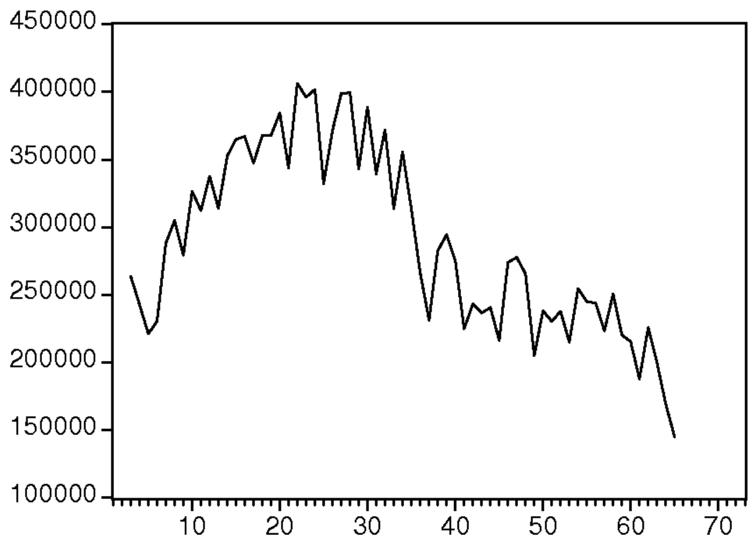


Figure 59: Number of passengers carried by United Airlines per quarter.

Dependent Variable: PAX5
Method: Least Squares
Date: 01/21/10 Time: 13:54
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 13 iterations
Backcast: 5 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	513864.2	76554.14	6.712428	0.0000
DUMMY	-58526.63	19200.79	-3.048137	0.0036
@TREND	-4370.729	1499.527	-2.914738	0.0052
AR(4)	0.689866	0.082843	8.327378	0.0000
MA(1)	0.522843	0.114180	4.579130	0.0000
MA(2)	0.614518	0.113511	5.413733	0.0000

R-squared	0.929813	Mean dependent var	292042.9
Adjusted R-squared	0.923192	S.D. dependent var	68000.47
S.E. of regression	18845.88	Akaike info criterion	22.62212
Sum squared resid	1.88E+10	Schwarz criterion	22.83340
Log likelihood	-661.3526	F-statistic	140.4250
Durbin-Watson stat	2.043954	Prob(F-statistic)	0.000000

Inverted AR Roots	.91	.00-.91i	.00+.91i	-.91
Inverted MA Roots	-.26-.74i	-.26+.74i		

Figure 60: United Airlines - regression on the quarterly number of passengers (filtered data).

Dependent Variable: PAX5
Method: Least Squares
Date: 01/21/10 Time: 14:00
Sample (adjusted): 6 65
Included observations: 60 after adjustments
Convergence achieved after 24 iterations
Backcast: 4 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	522589.0	42527.86	12.28816	0.0000
DUMMY	-69877.51	18709.15	-3.734938	0.0005
@TREND	-3810.315	991.5509	-3.842783	0.0003
AR(1)	-0.429398	0.097803	-4.390460	0.0001
SAR(4)	0.591879	0.081586	7.254673	0.0000
MA(1)	0.867856	0.126365	6.867870	0.0000
MA(2)	0.599171	0.102690	5.834748	0.0000

R-squared	0.941251	Mean dependent var	333356.2
Adjusted R-squared	0.934600	S.D. dependent var	78728.78
S.E. of regression	20133.58	Akaike info criterion	22.76745
Sum squared resid	2.15E+10	Schwarz criterion	23.01179
Log likelihood	-676.0234	F-statistic	141.5244
Durbin-Watson stat	1.942071	Prob(F-statistic)	0.000000

Inverted AR Roots	.88	-.00+.88i	-.00-.88i	-.43
	-.88			
Inverted MA Roots	-.43+.64i	-.43-.64i		

Figure 61: United Airlines - regression on the quarterly number of passengers (non-filtered data).

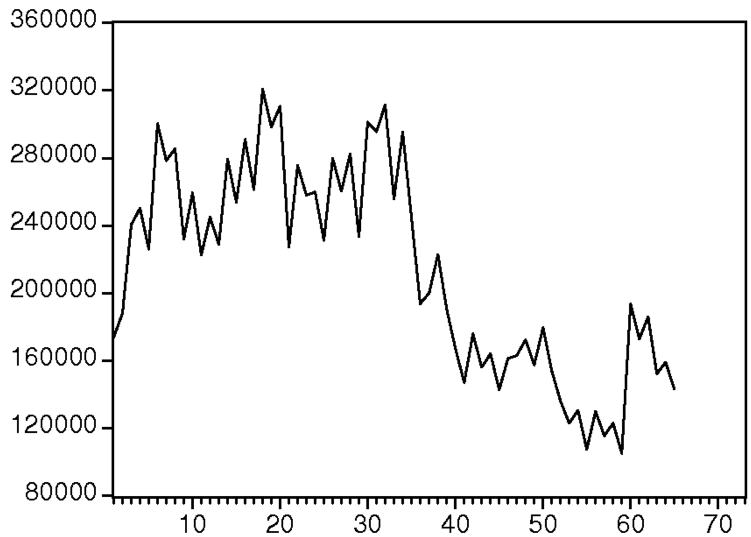


Figure 62: Number of passengers carried by US Airways per quarter.

Dependent Variable: PAX6
Method: Least Squares
Date: 01/21/10 Time: 14:40
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 12 iterations
Backcast: 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	267344.2	8060.288	33.16807	0.0000
DUMMY	-109693.0	10620.38	-10.32854	0.0000
AR(1)	-0.429863	0.092585	-4.642904	0.0000
AR(2)	0.556714	0.091936	6.055473	0.0000
MA(1)	0.997127	0.033607	29.66990	0.0000

R-squared	0.900232	Mean dependent var	214699.5
Adjusted R-squared	0.893352	S.D. dependent var	61796.14
S.E. of regression	20180.78	Akaike info criterion	22.73889
Sum squared resid	2.36E+10	Schwarz criterion	22.90898
Log likelihood	-711.2749	F-statistic	130.8378
Durbin-Watson stat	1.581042	Prob(F-statistic)	0.000000

Inverted AR Roots	.56	-.99
Inverted MA Roots	-1.00	

Figure 63: US Airways - regression on the quarterly number of passengers (filtered data).

Dependent Variable: PAX6
Method: Least Squares
Date: 01/21/10 Time: 14:52
Sample (adjusted): 6 65
Included observations: 60 after adjustments
Convergence achieved after 33 iterations
Backcast: 0 5

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	285182.8	10821.30	26.35385	0.0000
DUMMY	-126797.1	11166.48	-11.35515	0.0000
AR(1)	-0.819342	0.051731	-15.83851	0.0000
SAR(4)	0.743030	0.064832	11.46089	0.0000
MA(1)	1.543017	0.028907	53.37941	0.0000
MA(2)	0.969097	0.025384	38.17788	0.0000
SMA(4)	-0.899145	0.032669	-27.52285	0.0000

R-squared	0.922032	Mean dependent var	213485.1
Adjusted R-squared	0.913205	S.D. dependent var	63058.81
S.E. of regression	18577.73	Akaike info criterion	22.60660
Sum squared resid	1.83E+10	Schwarz criterion	22.85094
Log likelihood	-671.1979	F-statistic	104.4607
Durbin-Watson stat	1.890107	Prob(F-statistic)	0.000000

Inverted AR Roots	.93	.00-.93i		
Inverted MA Roots	.97	.00+.97i	-.00-.97i	-.77-.61i
	-.77+.61i	-.97		

Figure 64: US Airways - regression on the quarterly number of passengers (non-filtered data).

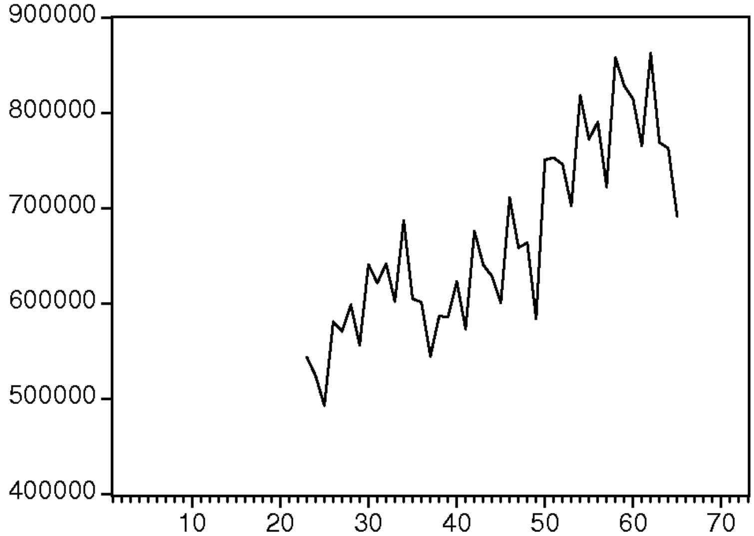


Figure 65: Number of passengers carried by Southwest Airlines per quarter.

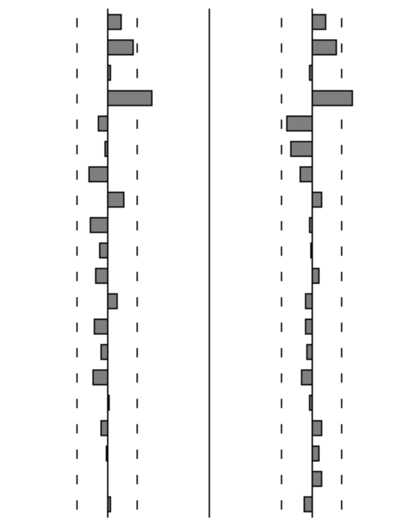
Dependent Variable: PAX7
Method: Least Squares
Date: 01/21/10 Time: 15:39
Sample (adjusted): 23 65
Included observations: 43 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	388125.5	28956.25	13.40386	0.0000
@TREND	6533.313	646.9963	10.09791	0.0000

R-squared	0.713222	Mean dependent var	669057.9
Adjusted R-squared	0.706228	S.D. dependent var	97138.40
S.E. of regression	52649.76	Akaike info criterion	24.62611
Sum squared resid	1.14E+11	Schwarz criterion	24.70802
Log likelihood	-527.4613	F-statistic	101.9679
Durbin-Watson stat	1.525955	Prob(F-statistic)	0.000000

Figure 66: Southwest Airlines - regression on the quarterly number of passengers (filtered data).

Date: 01/21/10 Time: 15:39
Sample: 23 65
Included observations: 43



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.143	0.143	0.9369	0.333
		2	0.261	0.246	4.1506	0.126
		3	0.033	-0.032	4.2044	0.240
		4	0.447	0.416	14.100	0.007
		5	-0.090	-0.251	14.511	0.013
		6	-0.032	-0.206	14.564	0.024
		7	-0.196	-0.118	16.631	0.020
		8	0.169	0.102	18.217	0.020
		9	-0.178	-0.030	20.022	0.018
		10	-0.076	-0.012	20.361	0.026
		11	-0.119	0.074	21.220	0.031
		12	0.093	-0.063	21.762	0.040
		13	-0.132	-0.068	22.886	0.043
		14	-0.072	-0.055	23.228	0.057
		15	-0.145	-0.103	24.677	0.054
		16	0.018	-0.027	24.701	0.075
		17	-0.066	0.103	25.025	0.094
		18	-0.008	0.071	25.029	0.124
		19	0.007	0.092	25.033	0.159
		20	0.036	-0.075	25.145	0.196

Figure 67: Southwest Airlines - correlograms of the residuals for the regression in Figure 66.

Dependent Variable: PAX7
Method: Least Squares
Date: 01/21/10 Time: 15:55
Sample (adjusted): 27 65
Included observations: 39 after adjustments
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	440934.0	96946.85	4.548203	0.0001
@TREND	5460.067	1823.321	2.994573	0.0049
AR(4)	0.623446	0.149988	4.156645	0.0002

R-squared	0.764005	Mean dependent var	682713.4
Adjusted R-squared	0.750894	S.D. dependent var	90907.93
S.E. of regression	45372.61	Akaike info criterion	24.35701
Sum squared resid	7.41E+10	Schwarz criterion	24.48497
Log likelihood	-471.9617	F-statistic	58.27279
Durbin-Watson stat	0.785936	Prob(F-statistic)	0.000000

Inverted AR Roots	.89			

Figure 68: Southwest Airlines - regression on the quarterly number of passengers (filtered data).

4.2.1 Summing up

Table 3 below summarizes the "best" models we get from our statistical regressions in Eviews6. We can again observe that there is no need for any carrier to use the differencing process to get the best model. The ARMA(0,2) model seems again to be adopted by 4 carriers, as it was the most adopted model for the regressions on fares, and for the same models: American Airlines, Continental Airlines, United Airlines. Here we found that the regressions for US Airways also need an autoregressive component to best fit the data. For Delta Air Lines and Northwest Airlines, we found an important autoregressive part in the model for the number of passengers. For Delta, this autoregressive component can also be related to the absence of seasonality coefficient in its regression, which leads to a stronger dependence from the closest past lags rather than from the usual quarterly number of passengers in a given quarter of the year. This may perhaps be explained by a higher volatility in the number of passengers carried by the airline.

We can see that 5 of the carriers underwent an important drop of their number of passengers after September 11, 2001, especially for Delta Air Lines and US Airways. The line "drop of number of passengers after period 34" (equal to the ratio between the dummy variable and the number of passengers carried during the period 34) shows the percentage decrease of the fares after period 34 which we can attribute to the consequences of the September 11, 2001 attacks. We can also observe that there seems to be a trend for the number of passengers flying with United Airlines to decrease by about 4370 passengers per quarter. On the contrary, as we saw for the fares, the low-cost carrier Southwest continuously increases its number of passengers by about 5460 per quarter, whereas the other airlines are stable or try to keep their passengers!

All these remarks should be qualified by the relative quality of these regressions. Even if these models are stationary and invertible, the low values of the Durbin-Watson show that it is difficult to find good models with no autocorrelation. We can partly explain this by the fact that, contrary to the fares, the number of passengers carried by an airline is more subject to volatility and depends

Carriers	**AA**	**CO**	**DL**	**NW**	**UA**	**US**	**WN**
use of filter on data	yes	no	yes	no	yes	yes	yes
AR degree	0	0	3	2	0	1	0
MA degree	2	2	0	1	2	2	0
differencing	no	no	no	no	no	no	no
ARIMA model	(0,0,2)	(0,0,2)	(3,0,0)	(2,0,1)	(0,0,2)	(1,0,2)	no
seasonal AR variable	yes	yes	no	yes	yes	yes	yes
seasonal MA variable	no	no	no	no	no	yes	no
trend variable	no	no	no	no	yes	no	yes
value of trend coefficient	no	no	no	no	-4370	no	5460
dummy variable at period 35	no	yes	yes	yes	yes	yes	no
value of dummy coefficient	no	-48700	-134600	-43100	-58500	-126800	no
passengers at period 34		233841	450327	246746	355741	295224	
drop of number of passengers after period 34		21%	30%	17%	16%	43%	
R^2	0.62	0.81	0.93	0.80	0.93	0.92	0.76
Schwarz criterion	23.2	22.1	23.2	22.5	22.8	22.9	24.5
Durbin-Watson statistic	1.77	1.41	1.79	1.80	2.04	1.89	0.79
Jarque-Bera statistic	0.77	2.07	1.54	9.83	0.10	0.31	3.10
F-statistic	31	59	183	44	140	104	58

Table 3: The carriers are reported with their IATA airline codes.

more on other economic, political and social parameters than the fares of a carrier.

4.2.2 Forecasting

With these 7 models, as for the fares, we could try to forecast the number of passengers for 1 on 2 years in the future. To ensure the reliability of our model, we can try to "forecast the past", as we did for the fares. We get the following graphs (see Figure 69), the real data plotted for all periods of observation (65 quarters), the best model for these data (dashed curve), and around this model, the possible distribution of the number of passengers at the 95% level of probability around the modelled value. We can see that, as expected, these models are not as good as the regressions for the fares of the airline, and hardly stabilize (especially for American Airlines). We also emphasize the small number of observations used for these regressions, which makes it very difficult to fit data correctly. However, we can again try to use this method to forecast the number of passengers after the 65 periods of observations. If we use the best model for each carrier and apply this model over 8 periods after period 65, we get the following graphs (Figure 70). We can also notice that the recent economic crisis also undergone by the U.S. airlines has led to a recent drop in the number of passengers for several airlines. As a consequence, our stationary models forecast for American Airlines, Continental Airlines, Delta and Northwest, an increase in passenger numbers for the next 2 years. On the contrary, United Airlines is here expected to continuously lose some passengers, following a global decreasing trend for the traffic of the carrier. For US Airways, this number of passengers may remain constant for the next 2 years, at around 150,000 passengers per quarter.

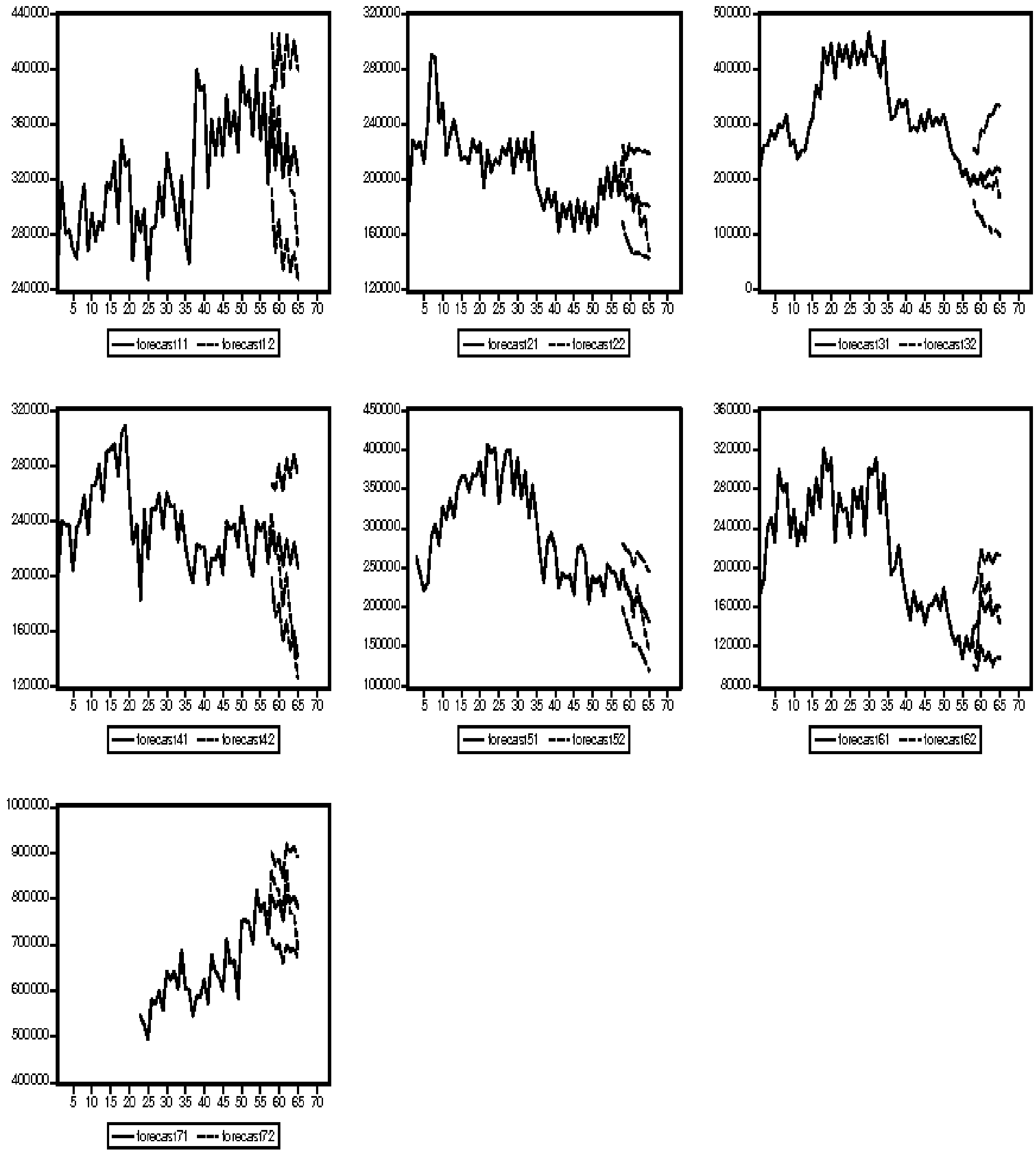


Figure 69: In-sample dynamic forecasts of the number of passengers carried from quarter 58 to quarter 65. Forecast 1 refers to American, forecast 2 refers to Continental, 3 refers to Delta, 4 refers to Northwest, 5 refers to United, 6 refers to US Airways and 7 refers to Southwest.

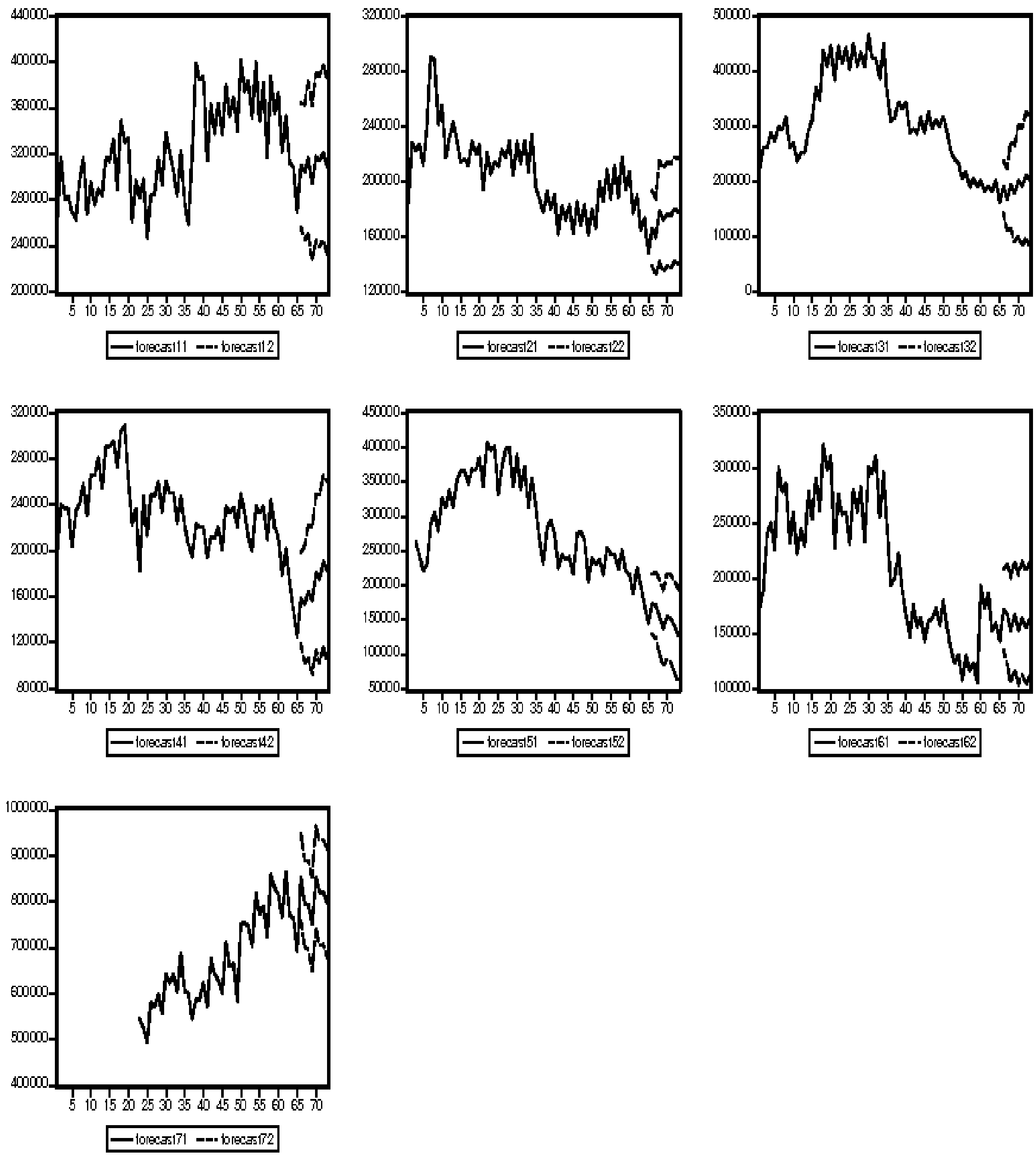


Figure 70: Out-of-sample dynamic forecasts of the number of passengers carried from quarter 65 to quarter 73. Forecast 1 refers to American, forecast 2 refers to Continental, 3 refers to Delta, 4 refers to Northwest, 5 refers to United, 6 refers to US Airways and 7 refers to Southwest.

5 Insertion of economic data

In order to improve the reliability of our models, we could then try to include some economic data in these regressions, especially for the regressions on fares.

5.1 Economic Data

5.1.1 Form 41 Financial Data

To implement real world economic data on the airlines into our models we use the so-called "Form 41 Financial Data" collected by the Bureau of Transport Statistics.[17] This dataset covers various financial and other core data of U.S. airlines with an annual operating revenue of more than $20 million. We use the "Schedule B-1" table out of which we only take the accrued salaries in every quarter, the "Schedule P-10" table containing information on the number of people employed by each airline in every year, the "Schedule P-12" table which gives quarterly information on different financial facts and the "Schedule P-12A" with data on each airline's monthly fuel consumption and expenses. In order to be able to properly include all financial information into the models, adjustments for the inflation effects have to be made. This is carried out as described in Section 3.1.3. The treatment of the raw data is carried out with Excel and the help of programs written in VBA which can be found in Appendix A.3.

5.1.2 B-1 Table Data

The only variable out of this table that we want to use is ACCR_SALARIES. Unfortunately for this variable there is no information available after 2006, and for Northwest even the 2006Q4 datapoint is missing. Delta Airline's graph, as seen in Figure 71, shows between 1997 and 2002 a very high level of salary expenses which seems odd and for which we cannot find an explanation. It can be

[17] The data is available on the Bureau's website: `http://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=135&Link=0`.

assumed though that such behaviour will not help when trying to model the development of Delta's fares using this variable.

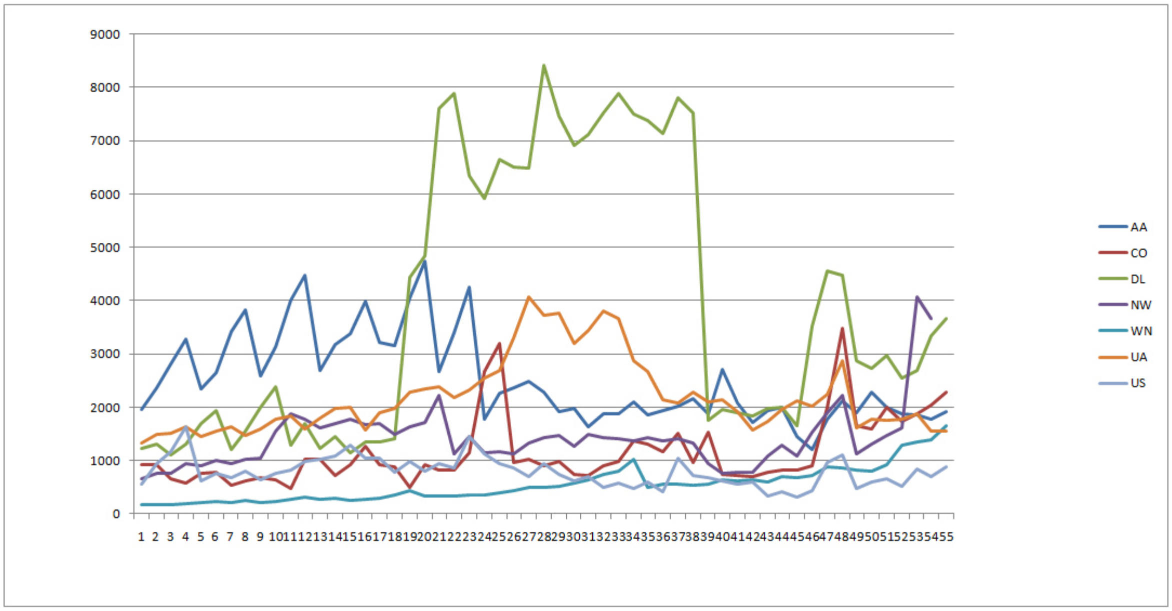


Figure 71: Accrued Salaries of the 7 airlines between 1993 and 2006, i.e. quarter no. 1 is 1993Q1 and quarter no. 55 is 2006Q4.

5.1.3 P-10 Table Data

Out of all the variables available in the table, we decided to download and examine the following:

- GENERAL_MANAGE
- PILOTS_COPILOTS
- OTHER_FLT_PERS
- PASSENGER_HANDLING
- TRANSPORT_RELATED

- TOTAL

The employee data was available for every year until 2008, so in order to get quarterly data that could be used in the model we use a simple linear regression between the figures for two years. Out of all the variables several turn out not to be usable. For GENERAL_MANAGE, i.e. the number of general management staff, the definition between the carriers seemed to vary considerably, so for example in 2005 Southwest Airlines reported 1809 general managers whereas US Airways reported only 12. Three other variables, OTHER_FLT_PERS, PASSENGER_HANDLING and TRANSPORT_RELATED were largely incomplete and thus we remove them as well. This leaves us with the following variables:

- **PILOTS_COPILOTS:** Figure 72 shows the number of pilots and co-pilots employed by the companies. As Southwest did not report any employee numbers in 1998 we approximated the missing figure through a simple linear regression using the figures for 1997 and 1999.

- **TOTAL:** This variable shows the total number of employees for each company. The data for Southwest in 1998 is missing here as well so we apply the same method as for the pilots to approximate a likely amount. Figures 73 and 74 show the variable's graph before and after the conversion into quarterly data.

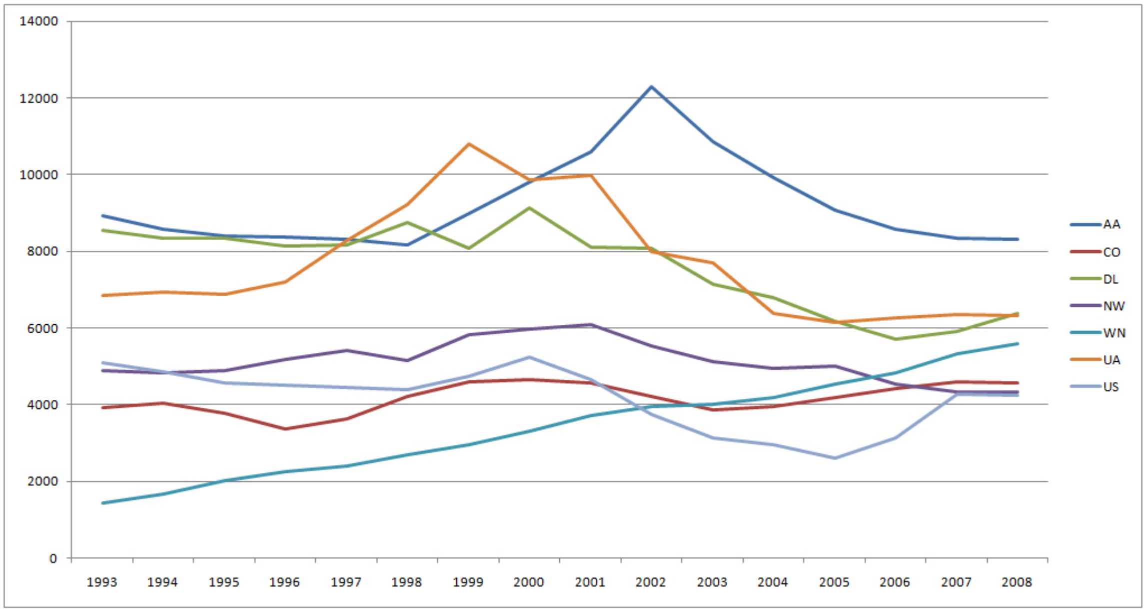


Figure 72: Number of pilots and co-pilots in the original yearly format.

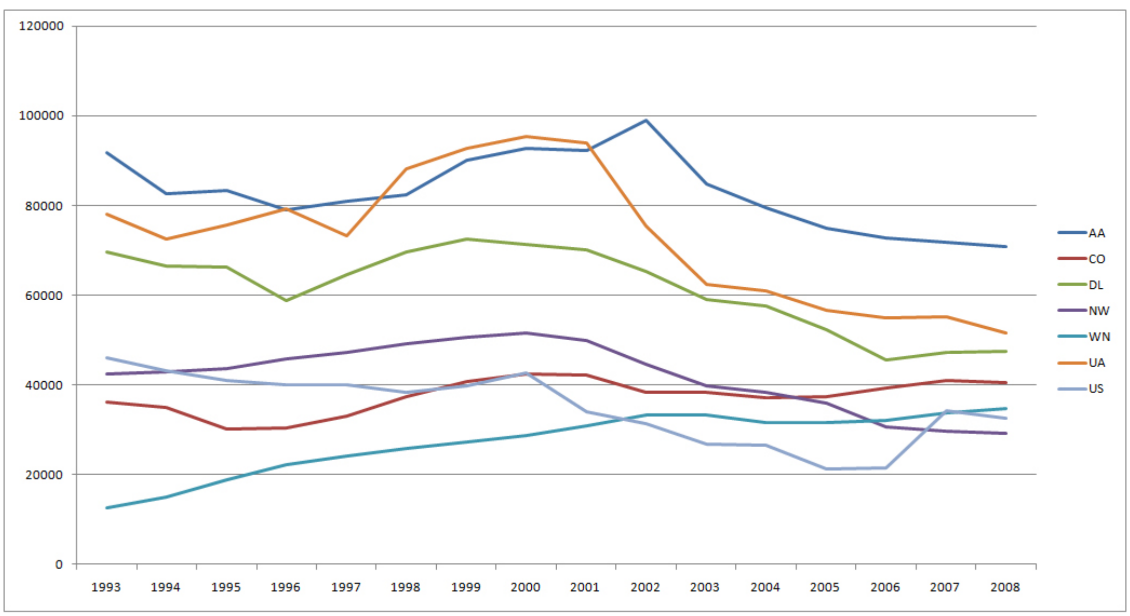


Figure 73: Total number of employees in the original yearly format.

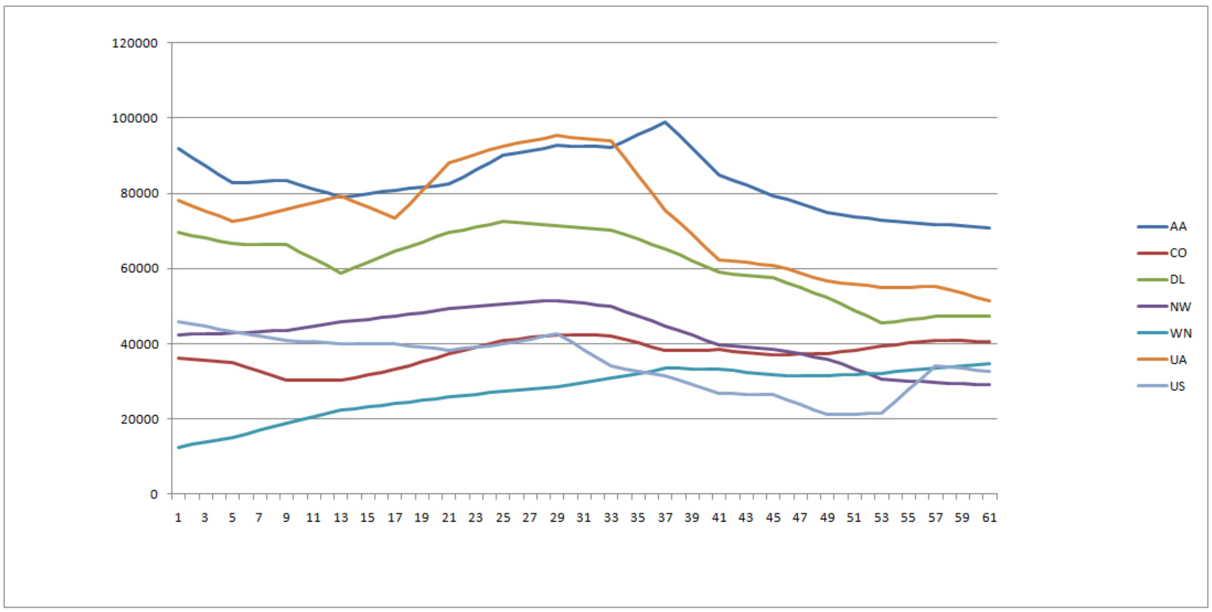


Figure 74: Total number of employees after the expansion to quarterly data through linear regression.

5.1.4 P-12 Table Data

Out of all the variables available in the table, we decided to download and examine the following:

- AIRCFT_SERVICES
- DEPREC_AMORT
- FLYING_OPS
- GENERAL_ADMIN
- GENERAL_SERVICES
- INCOME_PRE_TAX
- INCOME_TAX
- INTEREST_EXP_OTH
- INTEREST_LONG_DEBT
- MAINTENANCE
- NET_INCOME
- NON_OP_INCOME
- OP_EXPENSES
- OP_PROFIT_LOSS
- OP_REVENUES
- PAX_SERVICE

- PROMOTION_SALES
- PUB_SVC_REVENUE
- RES_CANCEL_FEES
- TRANS_REV_PAX
- TOTAL_MISC_REV
- TRANS_EXPENSES
- TRANS_REVENUE
- UNIQUE_CARRIER_NAME

All variables were reported separated into operational regions. Here, only the figures of the airline as a whole are interesting, and so we combine all regions to form a single datapoint per airline and quarter for every variable.

Several of the variables can at first glance be omitted as their content is obviously largely incomplete. This leaves us with the following variables:

- **AIRCFT_SERVICES:** This variable contains the airlines' expenses for aircraft and traffic servicing. The graphs in Figure 75 show a steady increase for most of the airlines most of the time. Only US Airways reports a longer period of decreasing aircraft servicing costs.

- **DEPREC_AMORT:** Depreciation and amortization of the airlines' assets. Figure 76 shows several peaks for various airline during the sample period, the reason for which remains unclear. These peaks could result from high one-time payments (or data errors) but a validation of these peaks certainly requires further research.

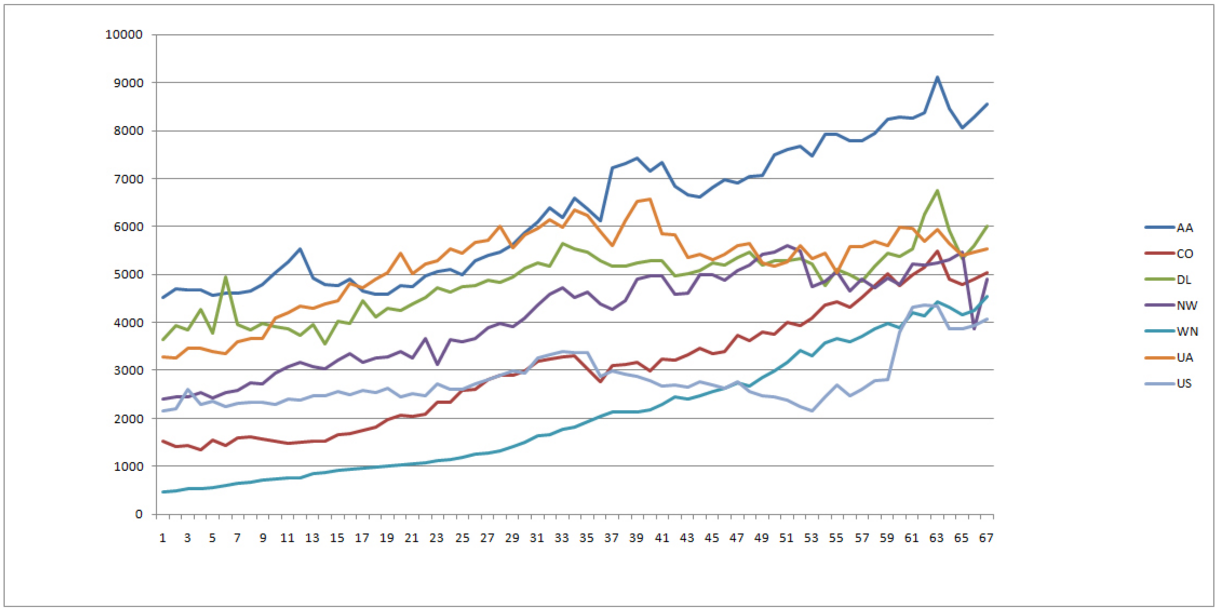


Figure 75: Aircraft and traffic servicing expenses.

- **FLYING_OPS:** The airlines expenses for their flight operations are contained here. The graph in Figure 77 shows an increasing increase of operating costs which drop sharply just before our observation period ends. This is probably due to the price for fuel which had a similar development in the last years and expenses for aircraft fuel are one of the major drivers of operations costs.

- **GENERAL_ADMIN:** GENERAL_ADMIN displays general and administrative expenses. Figure 78 shows several peaks that give the impression of being rather random and further analysis would have to be carried out to determine whether the peaks are real or errors. Especially the negative values found for US Airways and Continental in quarters 35 and 42 and the zero value for US in quarter 50 must raise doubts concerning the credibility of this statistic as expenses should normally not be negative. It is therefore also unlikely that this variable will be able to explain the development of the fares in our models.

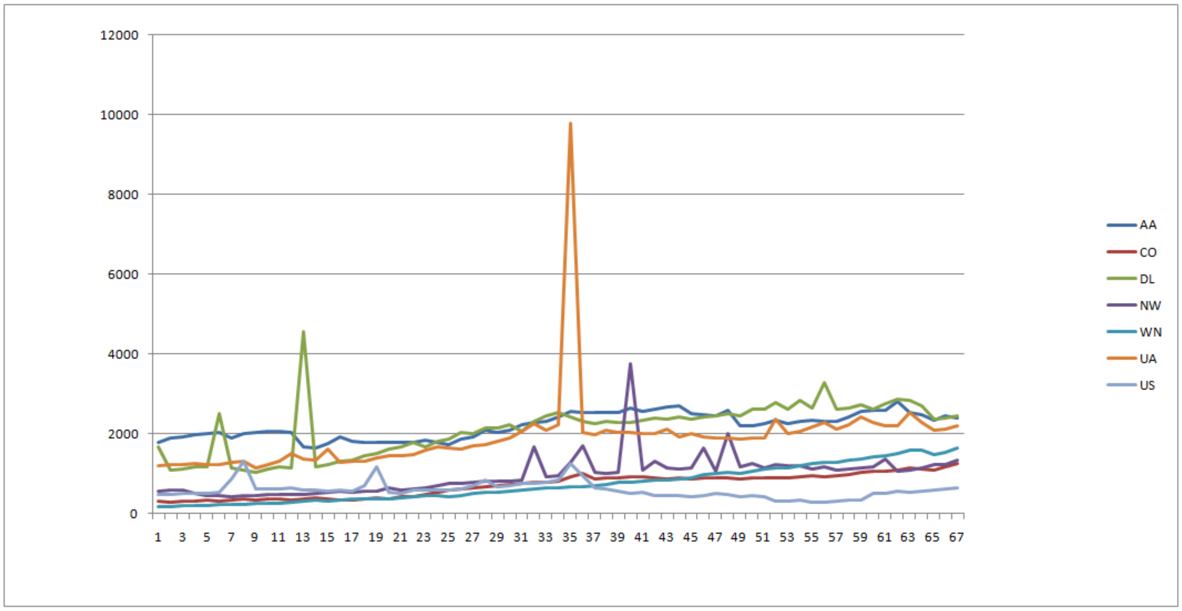


Figure 76: Expenses for depreciation and amortization.

- **GENERAL_SERVICES:** This variable will not undergo further examination since too much data is missing to form a dataset that could be used in the models.

- **INCOME_PRE_TAX:** The income before payment of income tax is depicted in Figure 79 and the corresponding variable. As with the administrative expenses variable, this variable also shows some very strong peaks. The most notable one is the low and high peak of the United Airlines data in 2005Q4 and 2006Q1 respectively. In February 2006, United Airlines left bankruptcy protection and the peaks are likely have something to do with that event. Two other peaks that are worth noting are those of Northwest and Delta in quarter 61 which almost coincides with the announcement of the merger of the two companies. All other peaks would require further research as we were not able to find any obvious explanations.

- **INCOME_TAX:** The INCOME_TAX variable shows the income tax paid by the airline in the respective period. The large number of peaks in the graphs in Figure 80 and the very

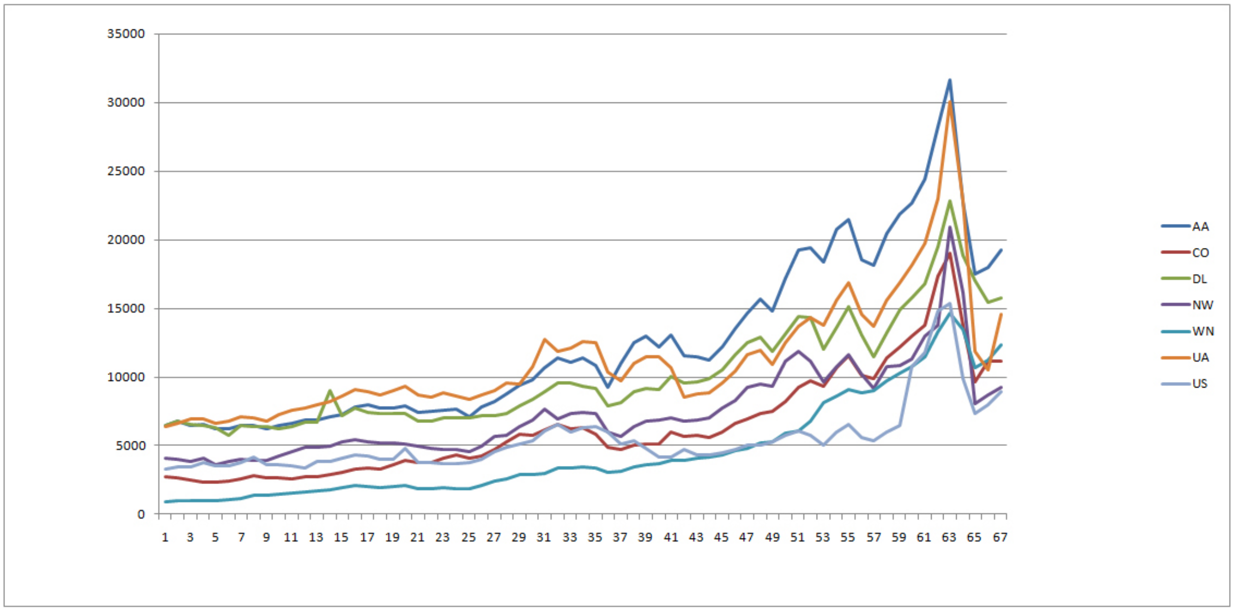


Figure 77: Flying operations expenses.

random appearance of the graphs indicate that the development of this variable will most likely not have a great explanatory effect on the fare models. Further research would be required though to identify whether the peaks are real or errors. The large number of gaps in the database for this variable will not help either.

- **INTEREST_EXP_OTH:** This variable (Figure 81) shows all interest expenses that are not interest expenses for long term debts. In general, if a company needs to borrow money on a short-term basis (which is usually more expensive than long term loans) this is not a good sign for the company's financial health. On the other hand, it is probably not a very important factor and thus, as with the income tax variable, due to the very high volatility of the development, explanatory power of this variable would be surprising.

- **INTEREST_LONG_DEBT:** Due to the expensive equipment (i.e. aircraft) needed in aviation, and due to the often low availability of cash money for purchases of aircraft, many

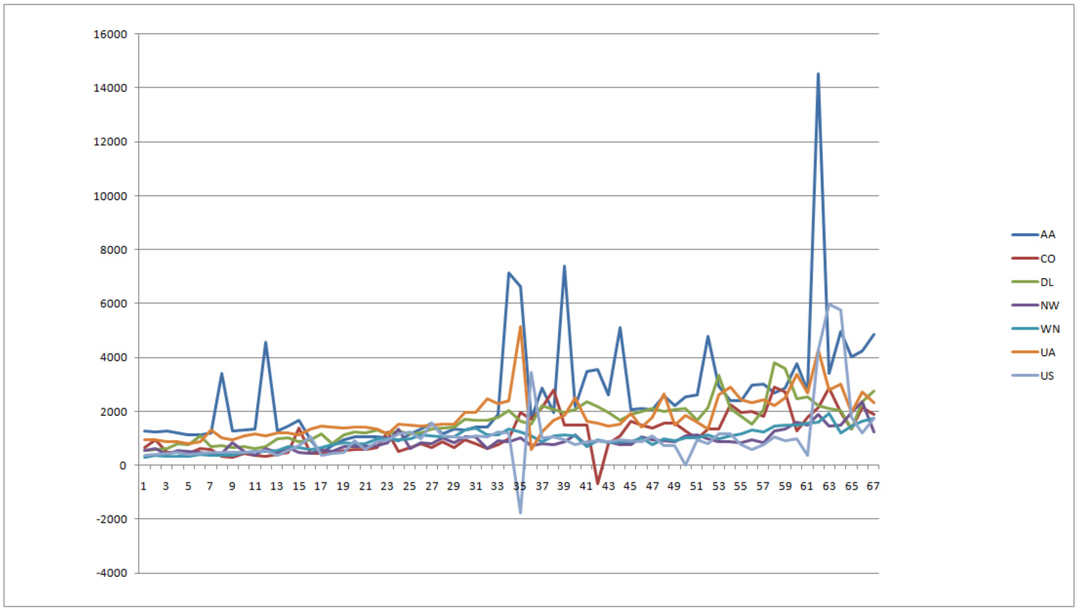


Figure 78: General and administrative expenses.

large carriers have to finance their planes through loans. Low-Cost carriers are often in a better situation and Southwest is no exception here. As one can see in Figure 82 Southwest clearly has to pay less interest than the other large carriers. The graphs feature a small number of peaks. Most of them require further research, only the US peak in quarter 55 is most probably a sign error as expenses should by definition be negative and the affected value would, multiplied by -1, fit into the graph very well.

- **MAINTENANCE:** This variable shows the expenses for aircraft maintenance.
- **NET_INCOME:** The net income graph (Figure 84) features similar characteristics to the pre tax income graph and has the same peak pattern which means that the UA double peak in 2005/2006 is likely a result of the end of the Chapter 11 bankruptcy protection for the carrier and the peak of DL and NW in quarter 61 coincides with the companies' merger.

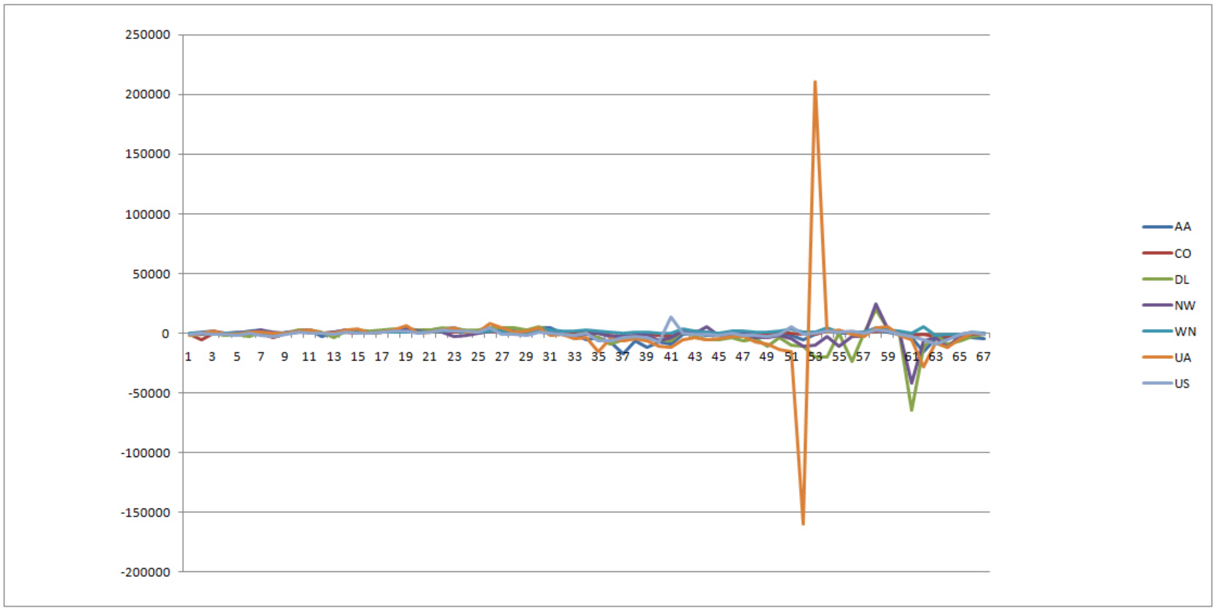


Figure 79: Income before income tax.

- **NON_OP_INCOME:** The non operating income visible in Figure 85 pretty much follows the net income and pre tax income graphs as well, which, together with the fact that the peaks do not appear in the OP_REVENUE and OP_PROFIT_LOSS variables (see Figures 88 and 87), supports the hypothesis that the peaks are related to accounting issues and financial business of the airline rather than having anything to do with the airline's actual operations.

- **OP_EXPENSES:** This variable seen in Figure 86 is similar to the graph in Figure 77, the effect of the oil price does not seem as severe, though, as in the flying operations graph because the OP_EXPENSES variable comprises more accounts than just those that directly deal with the flight operations.

- **OP_PROFIT_LOSS:** The operating profit and loss graph in Figure 87 features a clear seasonal effect which is introduced through the operating revenues, i.e. the revenues that with these carriers mainly come from the passengers whose numbers underly a seasonal periodicity.

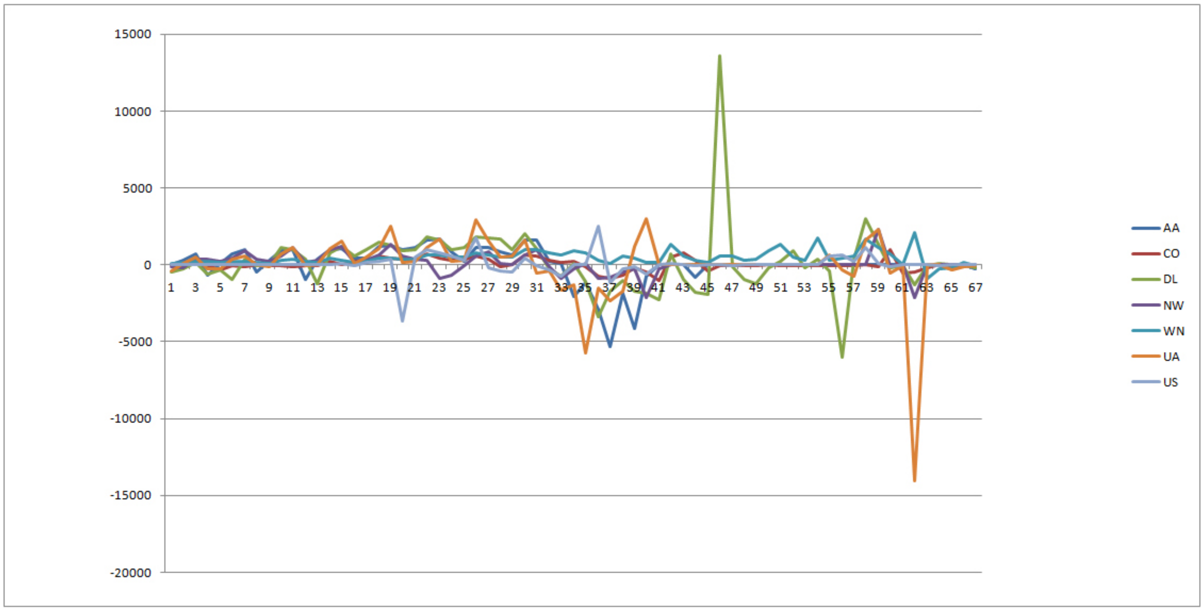


Figure 80: Income tax.

In this graph, certain effects in the airline economy like the September 11 attacks and the resulting losses and the losses due to high oil prices in 2008 are clearly visible.

- **OP_REVENUES:** The revenues of the airlines (Figure 88) are a direct result of the number of passengers flying and the fares these passengers are paying. Of course in the year 2001 effects of the terror attacks are visible, but also the financial crisis shock which reaches its peak in early 2009, after the high peak of the operating expenses due to high fuel prices which was in 2008.

- **PAX_SERVICE:** The airlines' expenses for passenger services seen in Figure 89 are far less volatile than other expenses, after 2001 they stayed relatively constant, despite increasing passenger numbers. The only exception is Southwest.

- **PROMOTION_SALES:** This variable shows the amount of money spent by the airline

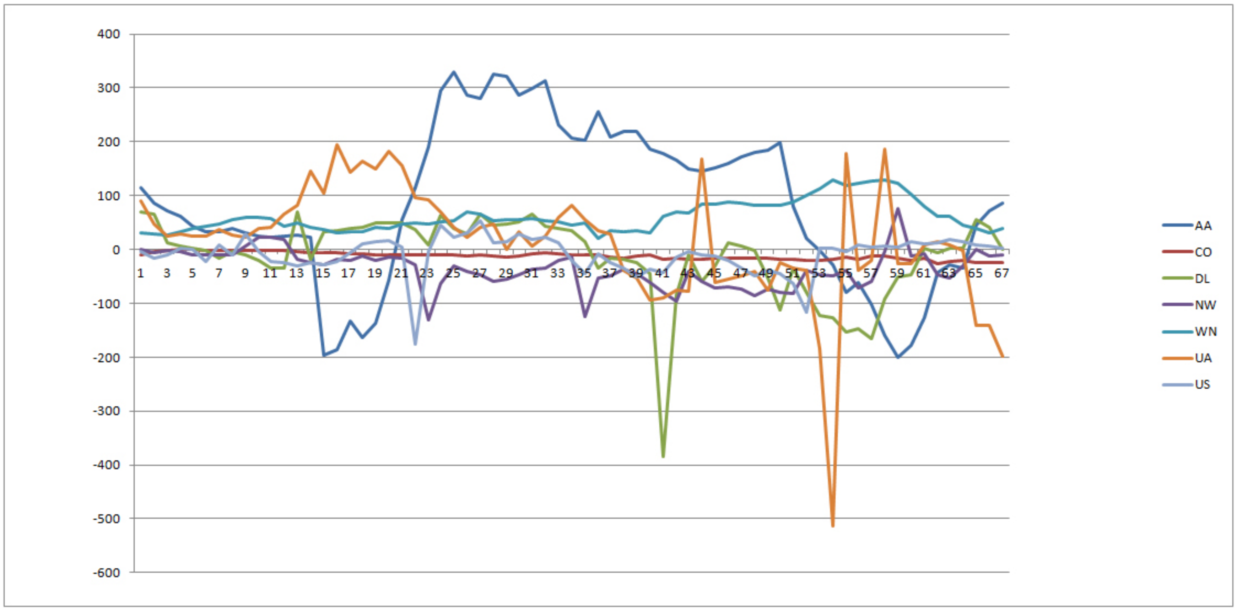


Figure 81: Other interest expense.

for promotion and sales. The graphs in Figure 90 suggest that the airlines have reduced their expenditure here in 2001, possibly due to their worsened financial situation and tighter budgets.

- **PUB_SVC_REVENUE, RES_CANCEL_FEES, TOTAL_MISC_REV, TRANS_REV_PAX:** These variables will not undergo further examination as too much data is missing to form a dataset that could be used in the models.

- **TRANS_EXPENSES:** All transport related expenses are included in this variable and it is therefore not surprising that its graph in Figure 92 is quite similar to the operating and flying operations expenses.

- **TRANS_REVENUE:** All transported related expenses are included in this variable. Only the Southwest figures seem a little small by a factor of possibly around 100, which looks like

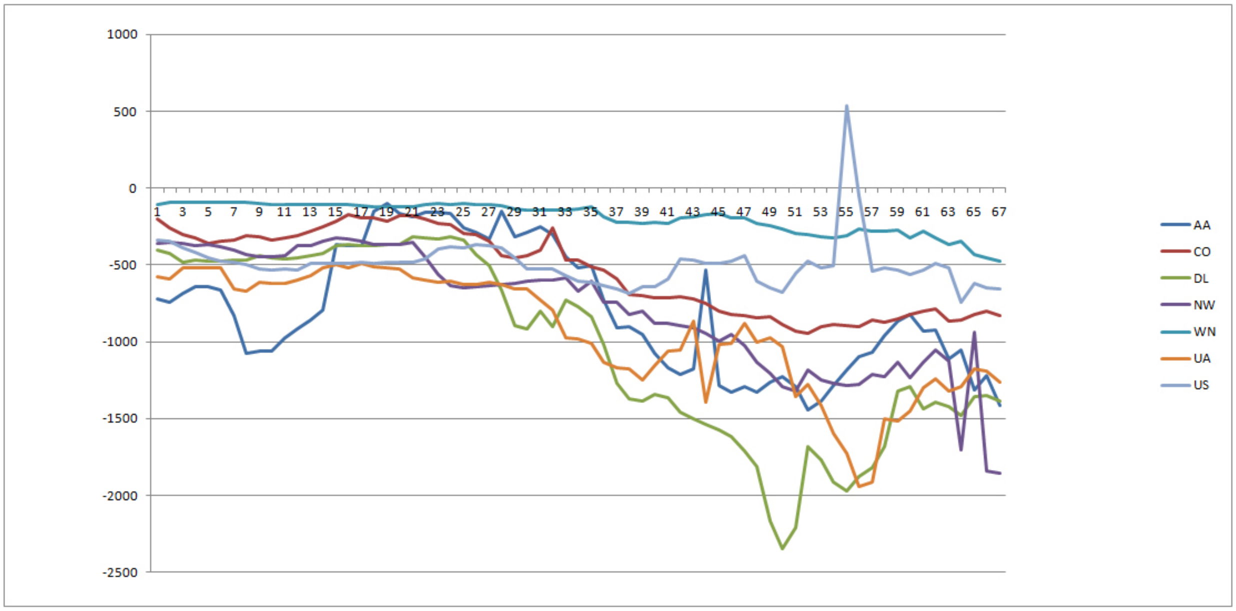


Figure 82: Interest on long-term debt and capital leases.

a reporting error. This should not have a very large effect on the models though. 2 minor corrections to the data have been made as well. Firstly, Southwest's figures for quarters 30 and 62 have been missing and were approximated through linear regression. Secondly, Delta Airlines reported a negative figure for quarter one, and as negative revenue would more or less mean that Delta paid its passengers for flying instead of getting paid, which seems unlikely, and since the absolute term of the figure fit very well to the figures in the quarter before and after we assumed that the mistake could be corrected by making the value positive.

- **UNIQUE_CARRIER_NAME:** This field identifies the airline that reported the number in the respective row of the database. In order to give every airline a unique name, numbers are added to the name should one name be used by different companies.

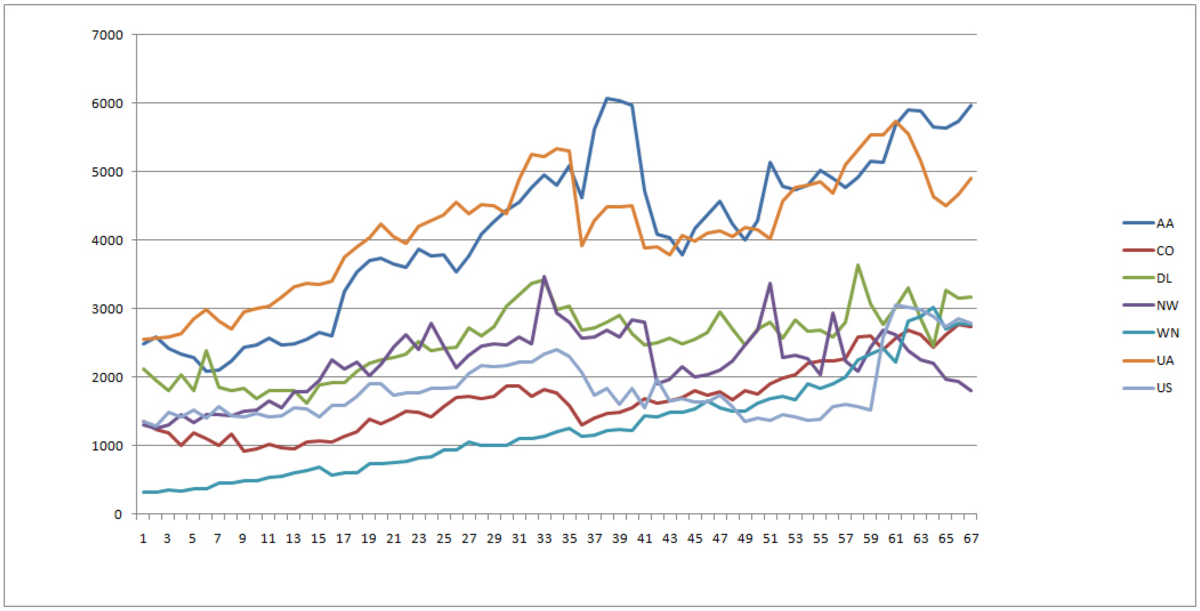


Figure 83: Aircraft maintenance.

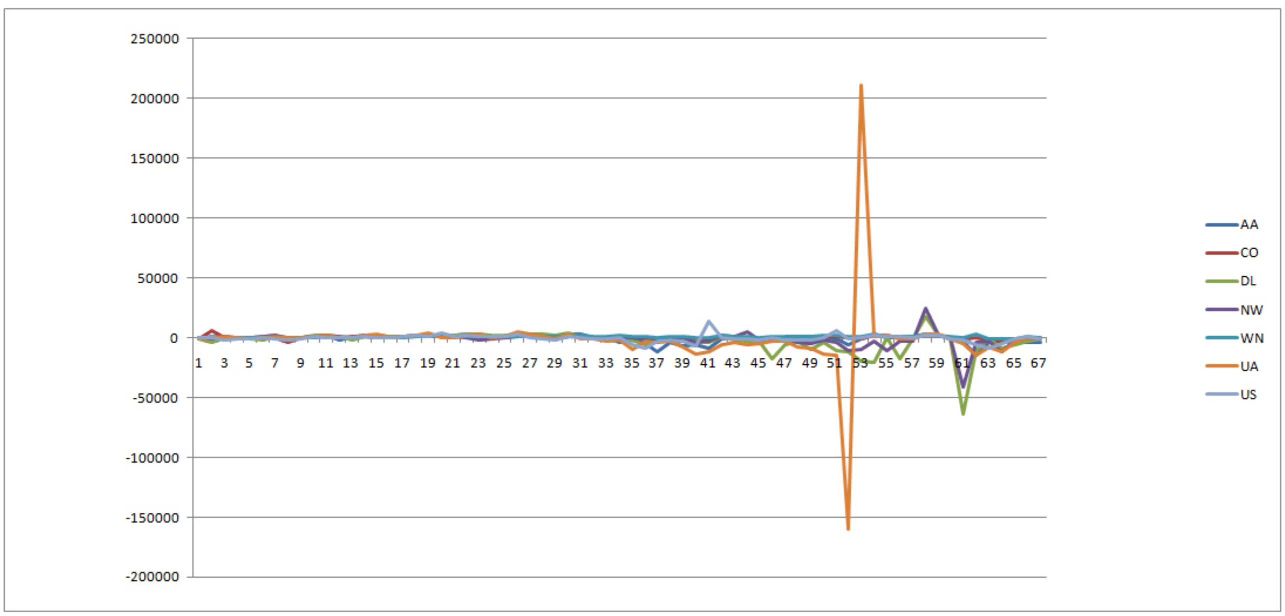


Figure 84: Net income.

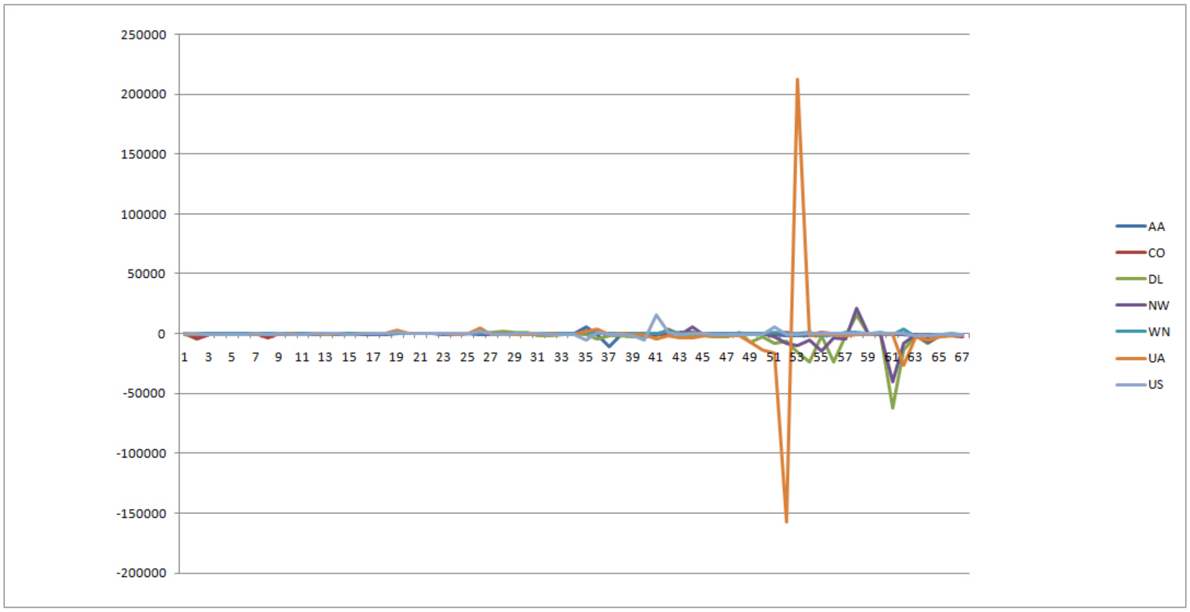


Figure 85: Nonoperating income and expenses.

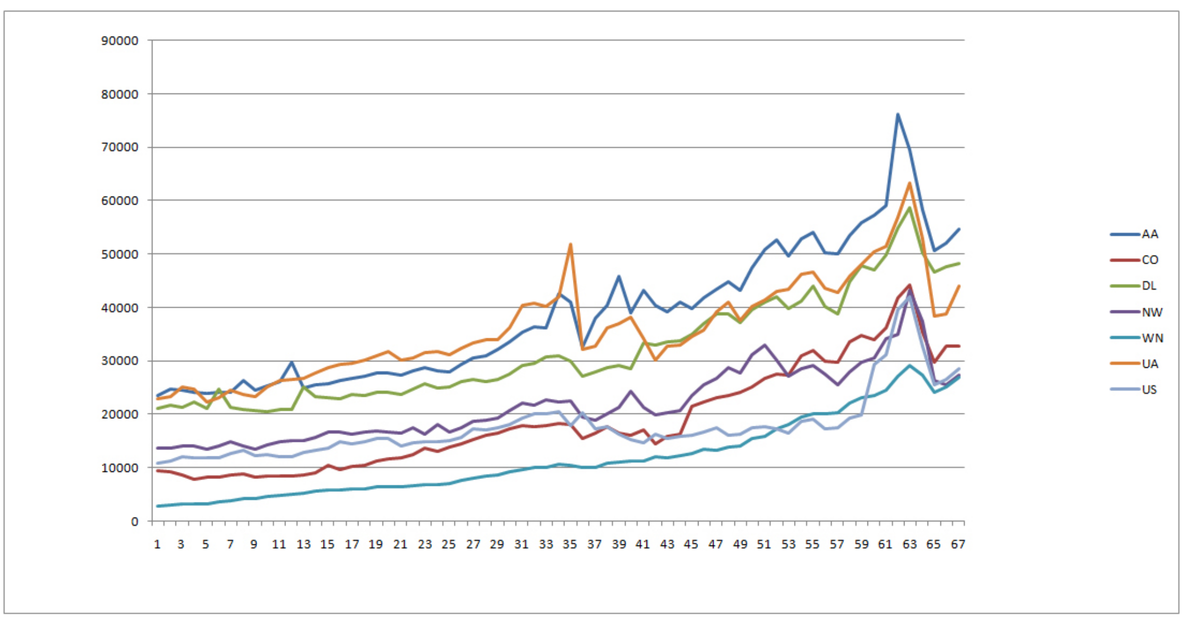


Figure 86: Total Operating Expenses.

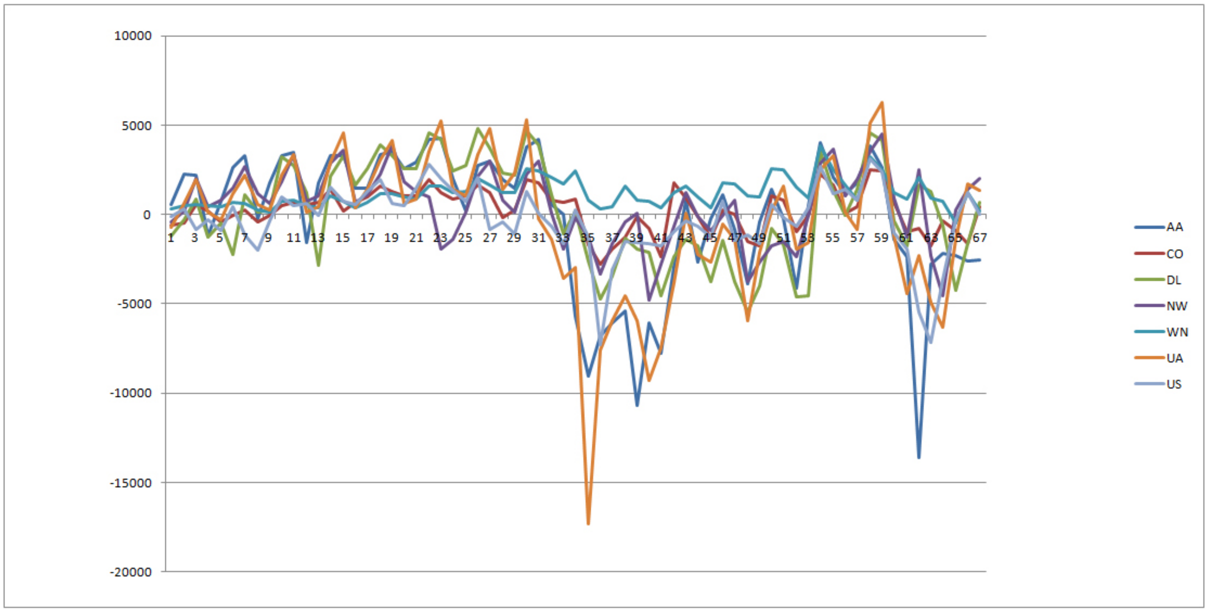


Figure 87: Operating profit or loss.

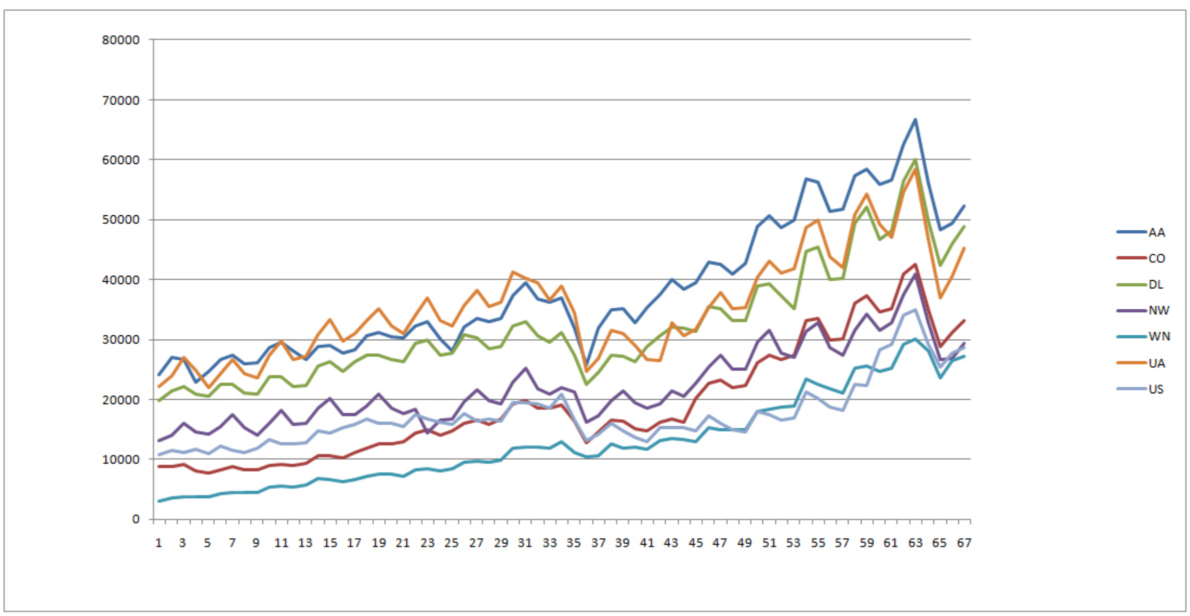


Figure 88: Total operating revenues.

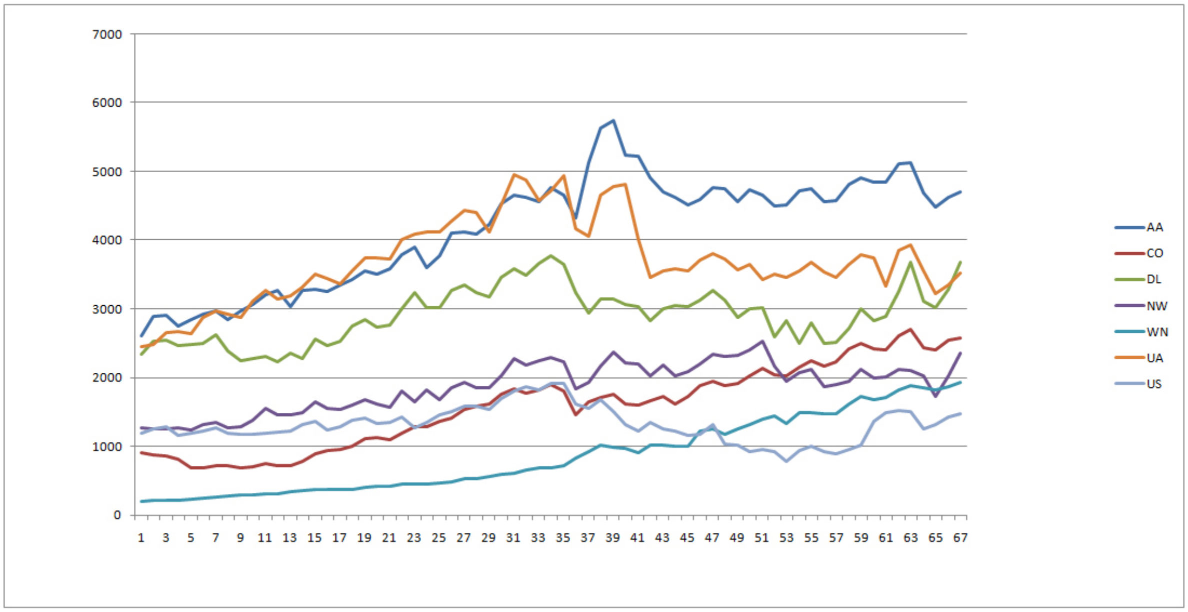


Figure 89: Passenger service.

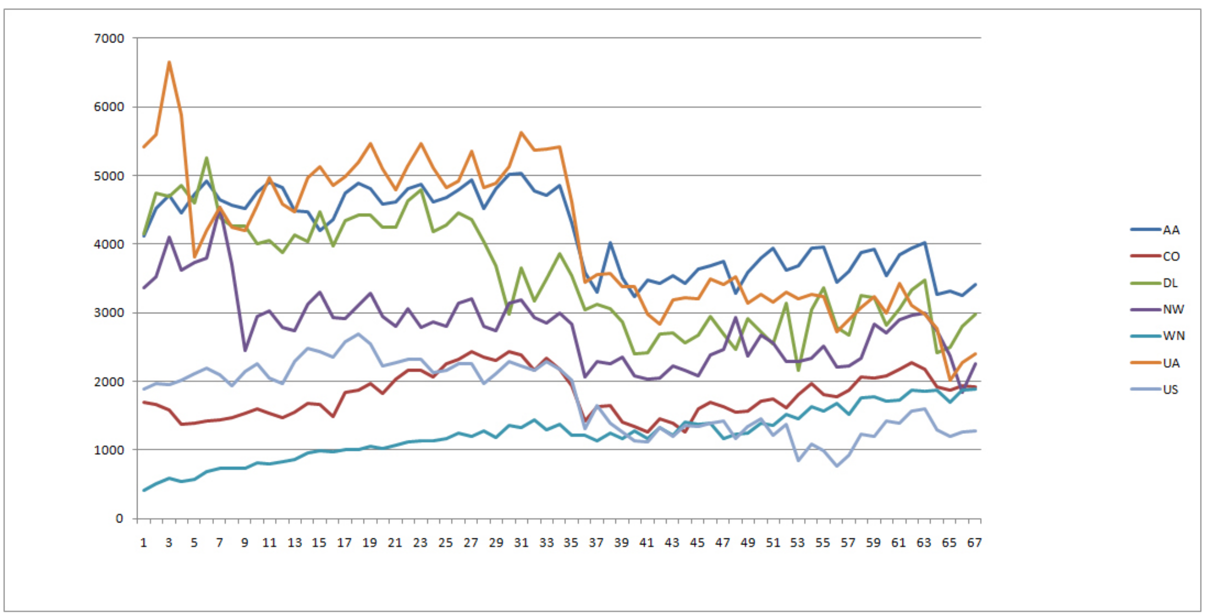


Figure 90: Expenses for promotion and sales.

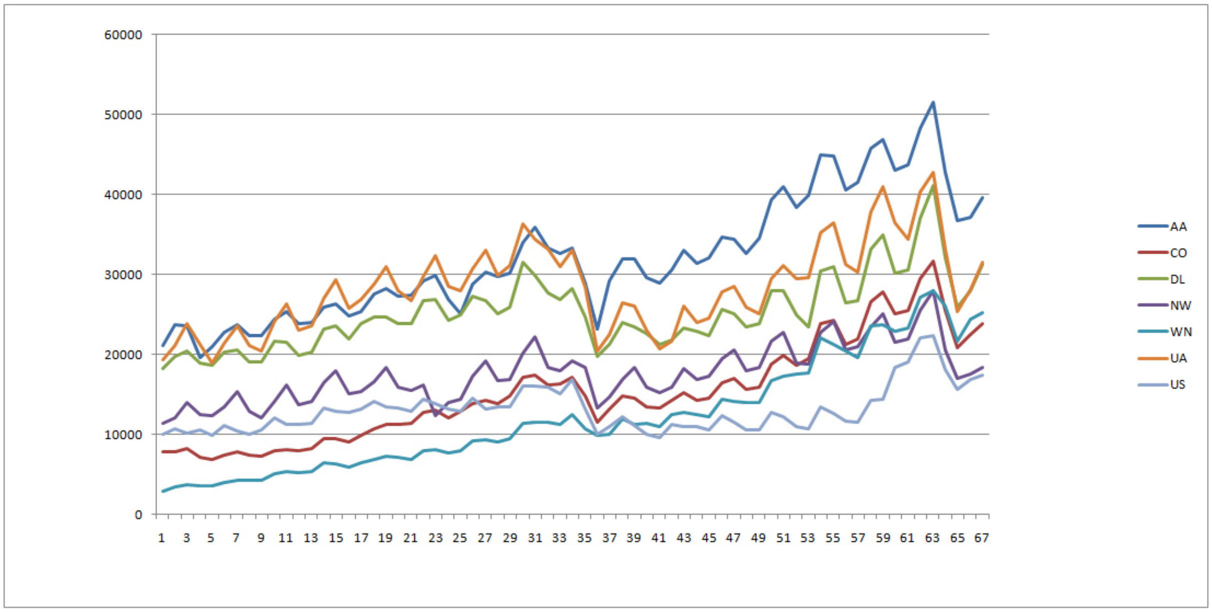


Figure 91: Passenger transport revenues.

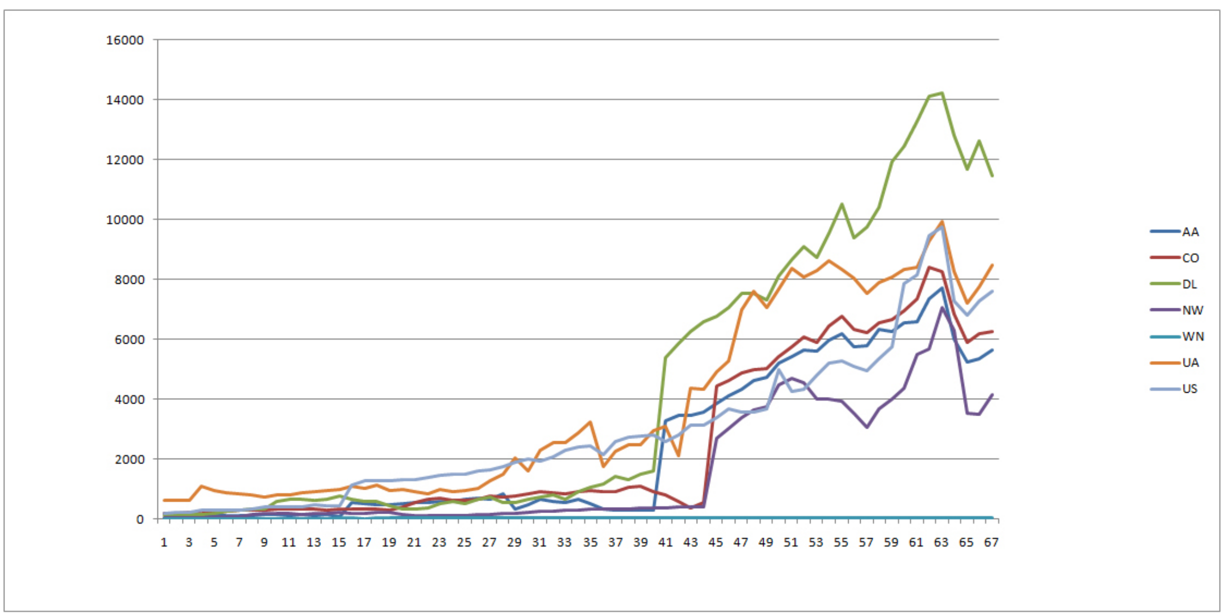


Figure 92: Transport related expenses.

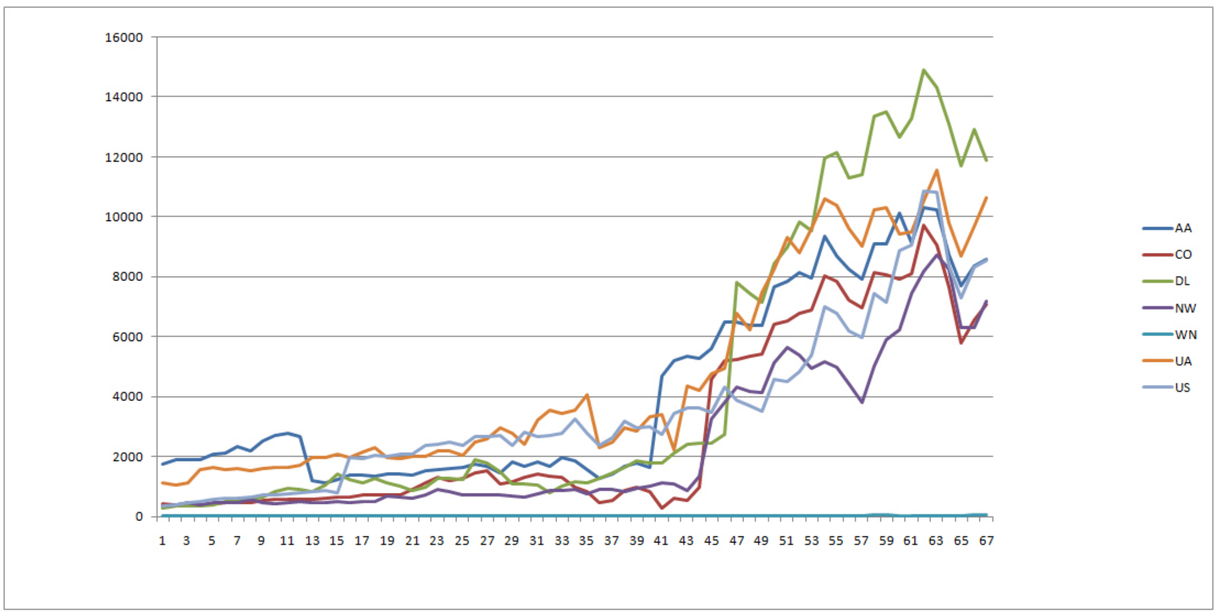


Figure 93: Transport related revenues.

5.1.5 P-12A Table Data

The P-12 table contains data about the airlines' total fuel consumption and their expenses for fuel. Detailed data for each airline is unfortunately only available for the years since 2000. From this database we downloaded the following four variables:

- **TDOMT_GALLONS:** This variable shows the fuel consumption for all domestic operations of an airline in gallons. The graph in Figure 94 shows the variable's graph.

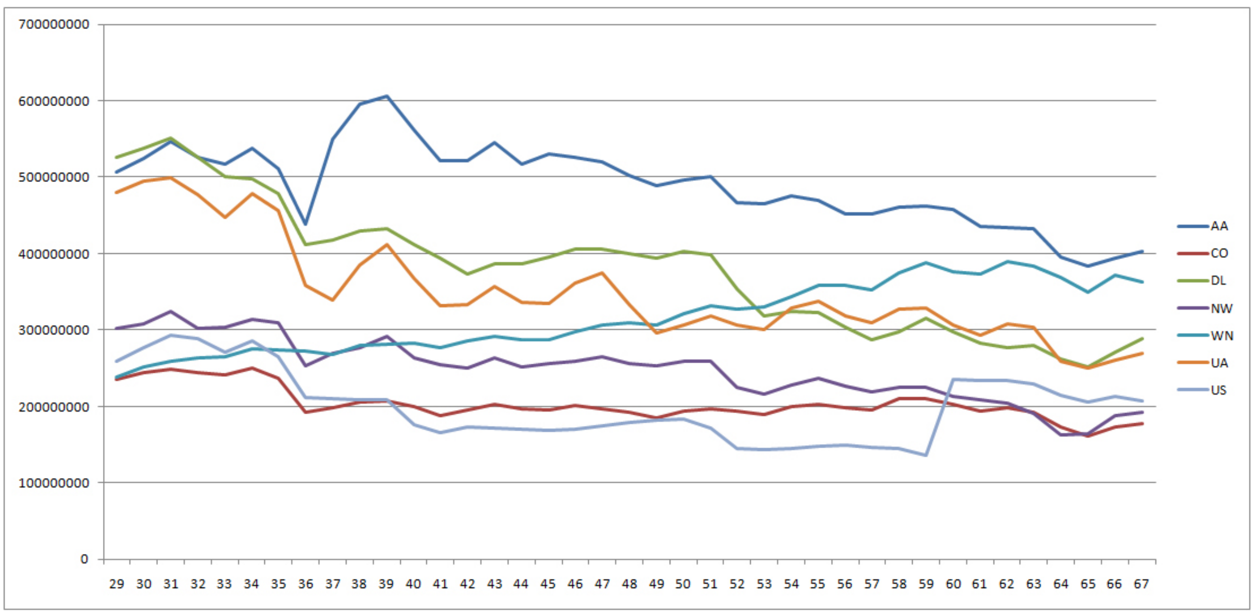


Figure 94: Domestic fuel consumption.

- **TOTAL_GALLONS:** This variable shows the fuel consumption for all domestic operations of an airline in gallons. The graph in Figure 95 shows the variable's graph.

- **TDOMT_COST:** This variable, graphed in Figure 96 shows the amount of money paid by each airline for its fuel used in its domestic operations. As the fuel price has a large impact on this variable, the oil price bubble in 2008 and the oil price drop in 2009 are clearly visible.

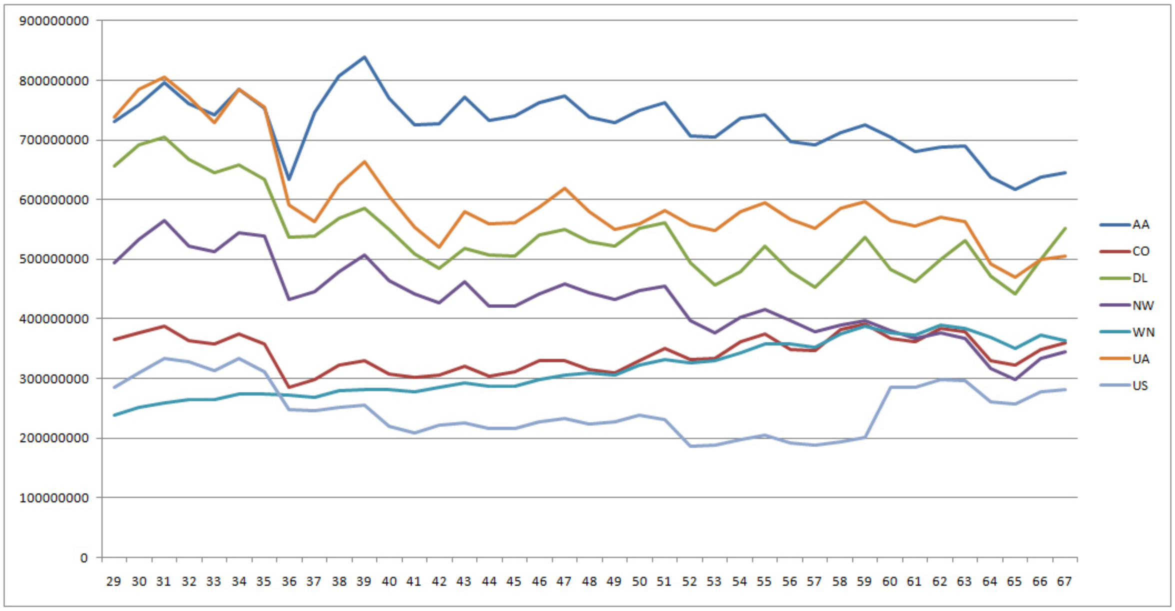


Figure 95: Total fuel consumption.

- **TOTAL_COST:** This variable, graphed in Figure 97 shows the amount of money paid by each airline for its fuel used in its domestic operations. As the fuel price has a large impact on this variable, the oil price bubble in 2008 and the oil price drop in 2009 are clearly visible.

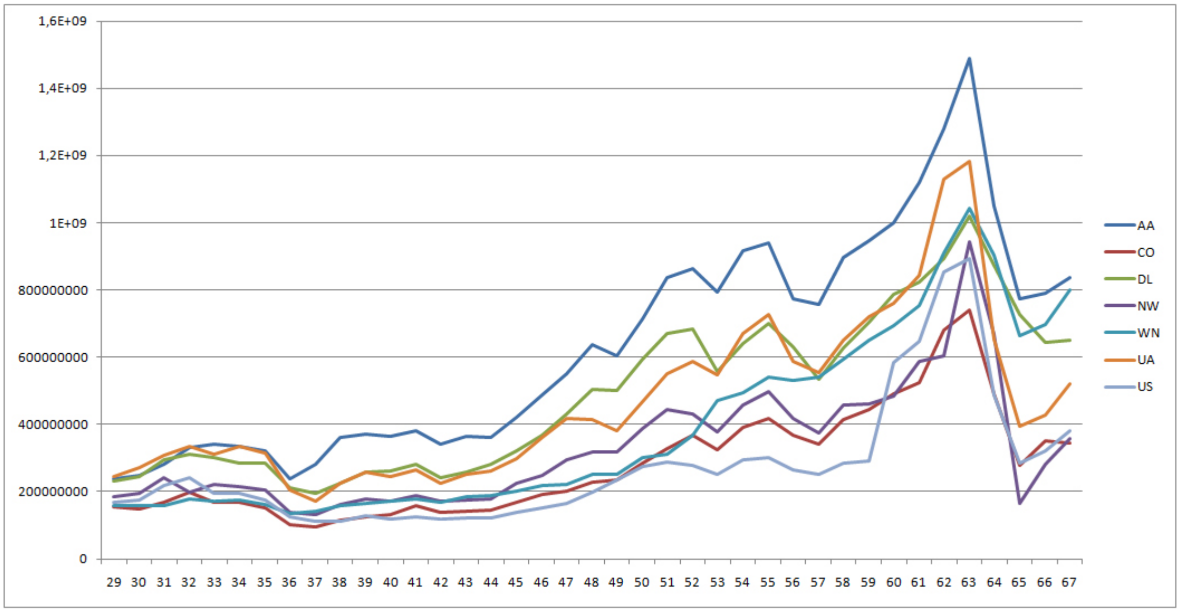


Figure 96: Domestic fuel expenses.

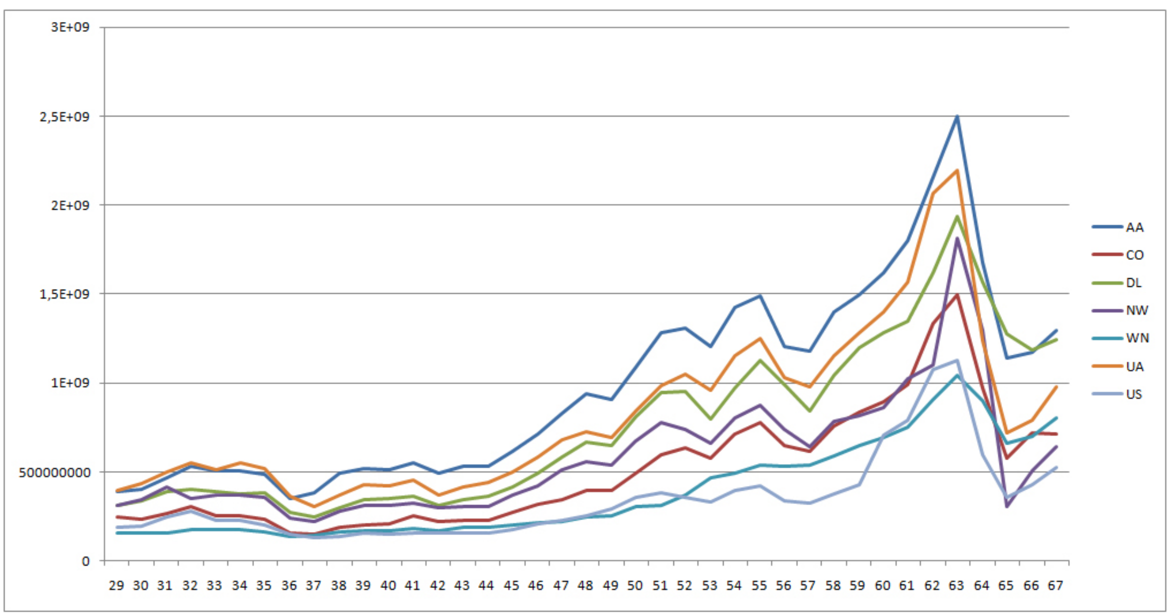


Figure 97: Total fuel expenses.

5.1.6 GDP

One of the most general macro economic variables we can have on the U.S. economy is the Gross Domestic Product of the country.[18] Consider for example the variation of the Gross Domestic Product (annual percent change) against time, between 1993 and 2008 (Figure 98). If we compare

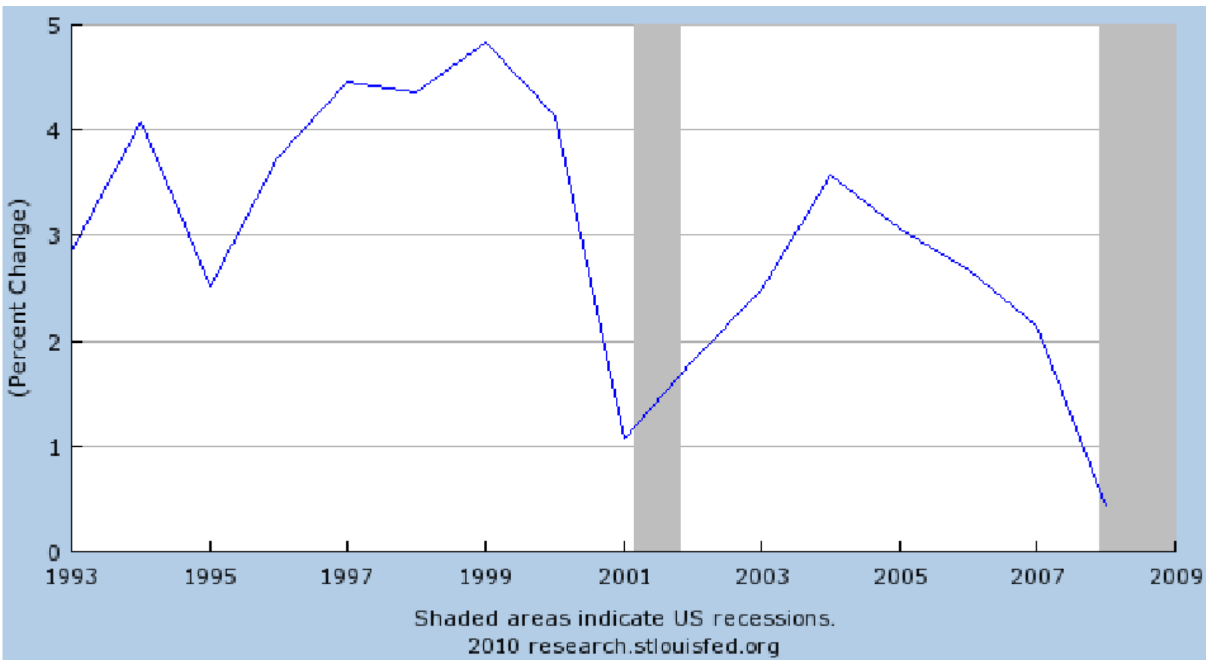


Figure 98: Real Gross Domestic Product (GDPCA) (% of annual variation against time).

this statistic with the variation of the fares of airlines over time (Figure 99), we can observe that the GDP and the airline average fares seem to be related: the same trends appear to connect the data, especially in the first periods of regression. Indeed, as the GDP statistic is the variation of the GDP from one year to another (and the GDP statistic being given at January the 1st), a spike at a given period means that the GDP had greatly increased during the year before this period. That can explain the apparent time lag between the variation of the GDP and the variation of the fares

[18]The GDP and other macroeconomic data are available on `http://research.stlouisfed.org`.

for the airlines, for the first periods of regression. On the contrary, this lag seems to be reversed in 2001 and for the periods of the U.S. terrorist attacks and the actual economic crisis. We can then assume that in 2001, the drop in GDP compared to the GDP in 2000 may be explained by the bust cycle of the dot-com bubble which happened during that year. The U.S. recession we can observe in Figure 98 can then be a combination of the end of the dot-com bubble and the terrorist events of September 11. In 2007, we can also see the beginning of the financial crisis, which has become the current economic crisis where the airlines, as many other U.S. industries, have faced financial problems.

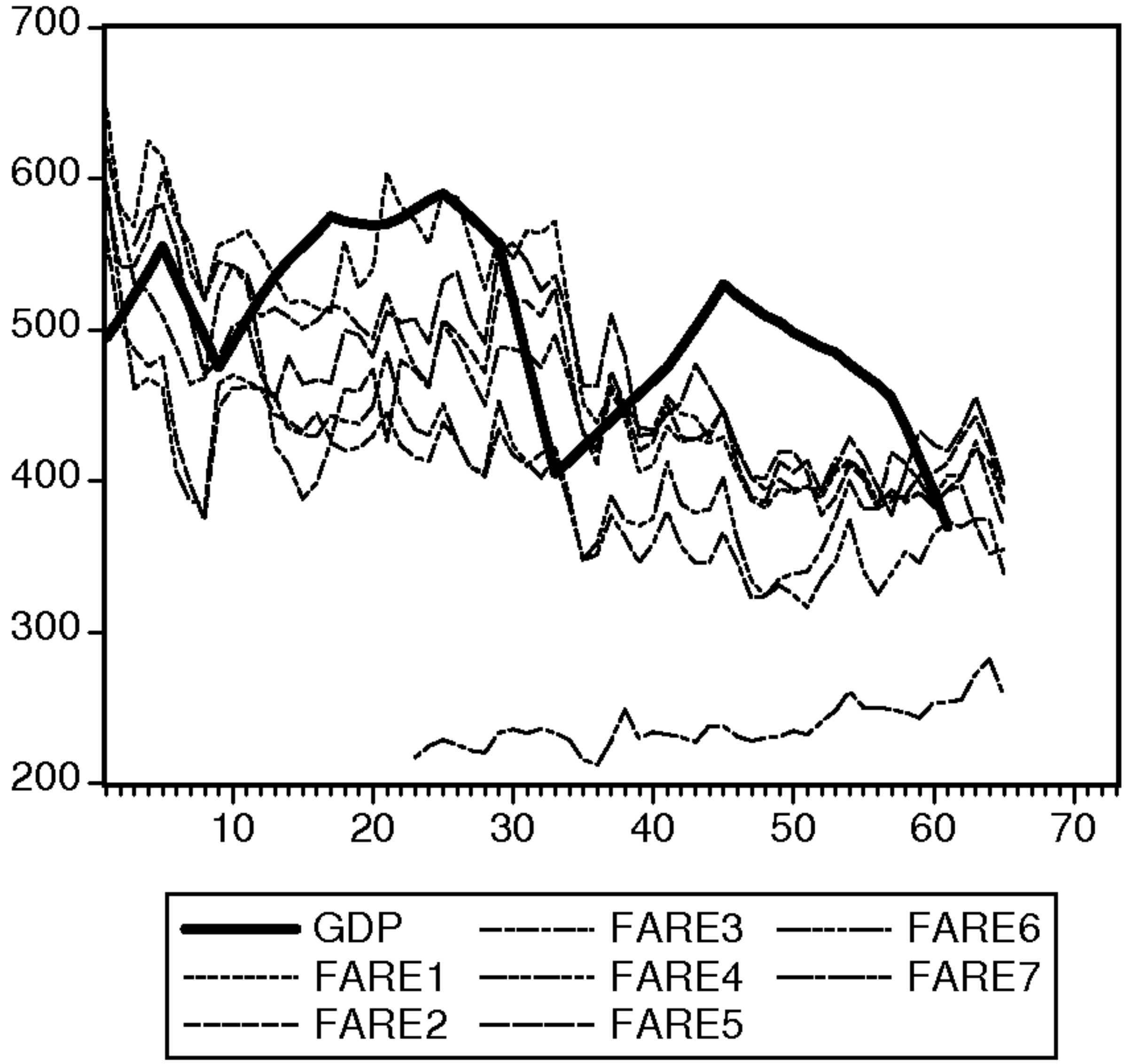


Figure 99: GDPCA (scaled to appear with the average fares of the airlines), and the quarterly average fares of the airlines (fare1 is for American, 2 is for Continental, 3 for Delta, 4 for Northwest, 5 for United, 6 for US Airways, 7 for Southwest).

5.2 Insertion of economic data into ARIMA models for fares

We can notice that in our ARIMA models, the use of dummy variables or trend was a tool to proxyh macroeconomic variables in our regressions. In fact, we can assume that these 2 effects can also be found in the economic data of the carrier, such as its operational profits, revenues, expenses, the price and consumption of fuel, the net income.

If we run some regressions using our best models for average fares (as computed above) and including economic data such as the fuel, we can observe that models for the fares deteriorate: many variables become insignificant, and not necessarily the economic data. In fact, we can try to explain these results by the fact that the dummy variable or trend variable also try to take into account macroeconomic trends, thus there may be some autocorrelation between these 2 variables and the economic data we want to include.

Now if we run these regressions using our best models for fares without the dummy and trend variables, we can have interesting results. As for the ARIMA models, we decided to run all the possible regressions (using our best ARIMA model with one economic variable), and retained the best model for each carrier according to the Schwarz Information Criterion.

- **American:** For American Airlines, we observe that all the regressions have insignificant coefficients if we include the trend variable in the model. If not, then we can observe that the best model we can get with EViews6 is the model with no trend variable and the variable of operational expenses (Figure 100). The R^2 statistic is very high for this regression (0.93), all the coefficients are individually significant, the model is stable and not autocorelated. Thus we can infer that the fare setting for American Airlines is highly correlated with the operational expenses of the carrier, which seems to be quite logical. If we consider the variable of operational revenues, transport expenses or transport revenues instead of this variable, the result is very similar and confirms this trend for the airline to determine its average fares with the operational expenses.

Dependent Variable: FARE1
Method: Least Squares
Date: 01/24/10 Time: 22:49
Sample (adjusted): 1 65
Included observations: 65 after adjustments
Convergence achieved after 12 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO1_13	-0.001709	0.000579	-2.949296	0.0045
C	600.4467	19.10238	31.43309	0.0000
DUMMY	-106.8731	15.57458	-6.862022	0.0000
MA(1)	0.896658	0.089225	10.04943	0.0000
MA(2)	0.385501	0.084943	4.538349	0.0000

R-squared	0.934951	Mean dependent var	488.0287
Adjusted R-squared	0.930615	S.D. dependent var	80.47577
S.E. of regression	21.19818	Akaike info criterion	9.019511
Sum squared resid	26961.77	Schwarz criterion	9.186772
Log likelihood	-288.1341	F-statistic	215.5967
Durbin-Watson stat	2.239931	Prob(F-statistic)	0.000000
Inverted MA Roots	-.45+.43i	-.45-.43i	

Figure 100: American Airlines - regression on quarterly average fares. eco1_13 stands for the variable of Operational Expenses.

- **Continental:** For Continental Airlines, EViews6 gives as the best regression for fares (Figure 101) the model which includes the variable of the total number of employees in the airline! It appears to be quite unlikely that the average fares of an airline can be connected to the total number of employees. However, we could imagine that if the number of employees increases, it is an additional expense for the airline, and as these new employees may improve the quality of the service offered to the passenger, this latter would consequently be less reluctant to pay more if there are more people taking care of him during his travel!

  If we do not include the dummy variable in our regression, then we get the variable of transport expenses as the best variable explaining the data of fares for the carrier (Figure 102). As for Continental Airlines, we can observe that the sign of the economic coefficient is negative, which would mean that the greater the transport expenses are, the lower the fares

Dependent Variable: FARE2
Method: Least Squares
Date: 01/24/10 Time: 23:09
Sample (adjusted): 1 61
Included observations: 61 after adjustments
Convergence achieved after 11 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO2_21	0.004463	0.001923	2.320354	0.0240
C	303.7066	70.70151	4.295617	0.0001
DUMMY	-61.54282	13.52376	-4.550719	0.0000
MA(1)	0.889295	0.080480	11.04982	0.0000
MA(2)	0.513502	0.088953	5.772743	0.0000

R-squared	0.756709	Mean dependent var	446.2246
Adjusted R-squared	0.739331	S.D. dependent var	42.35422
S.E. of regression	21.62428	Akaike info criterion	9.063923
Sum squared resid	26186.14	Schwarz criterion	9.236946
Log likelihood	-271.4497	F-statistic	43.54417
Durbin-Watson stat	2.405410	Prob(F-statistic)	0.000000
Inverted MA Roots	-.44-.56i	-.44+.56i	

Figure 101: Continental Airlines - regression on quarterly average fares. eco2_21 stands for the variable of the total number of employees.

become. The sign does not change if we include the transport revenues variable instead of the first one (as the transport expenses and revenues will not be so different for given airline). Indeed, we can assume that the transport revenues or expenses are a sign of the strength and development of the airline, and a larger airline (in terms of demand) will maybe have more ability to set lower fares than a startup company.

- **Delta:** For Delta Air Lines, the best model we get is the model with the variable of Passenger Services (Figure 103). This variable is indeed similar to the variable Operational Expenses, as the cost of the passenger services is a component of the operational expenses. So the sign of the coefficient for this regression (negative) is coherent with the above explanations for the model of Continental Airlines. Here again, the R^2 statistic (0.91) is very high, and our regression consequently seems to explain a important part of the fares.

Dependent Variable: FARE2
Method: Least Squares
Date: 01/24/10 Time: 23:10
Sample (adjusted): 1 65
Included observations: 65 after adjustments
Convergence achieved after 11 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO2_24	-0.007320	0.002468	-2.966053	0.0043
C	459.3858	9.377713	48.98698	0.0000
MA(1)	0.927830	0.076417	12.14171	0.0000
MA(2)	0.550311	0.083340	6.603195	0.0000

R-squared	0.706116	Mean dependent var	444.5904
Adjusted R-squared	0.691663	S.D. dependent var	41.83435
S.E. of regression	23.22983	Akaike info criterion	9.188315
Sum squared resid	32917.12	Schwarz criterion	9.322123
Log likelihood	-294.6202	F-statistic	48.85500
Durbin-Watson stat	2.174646	Prob(F-statistic)	0.000000
Inverted MA Roots	-.46-.58i	-.46+.58i	

Figure 102: Continental Airlines - regression on quarterly average fares. eco2_24 stands for the variable of Transport Expenses.

- **Northwest:** For Northwest, we get very different behaviour than for the other airlines, as we can observe that it is difficult to find a reliable and coherent model for the fares if we do not include the trend variable in the regression. This may be explained by the difficulties of the airline and the fact that it has recently been forced to merge with Delta Air Lines.

  If we consider the entire model with the trend variable, EViews6 gives the model with the Flying Operations as the best regression (Figure 104). All the coefficients are here again significant at the 95% level, and the regression explains more than 96% of the fare data according to the R^2 statistic.

- **United:** For United Airlines, again the best economic variable we may include in the regression seems to be the Passenger Services. The sign of the coefficient is negative as we found for the regression for Delta, and the coefficient of determination R^2 is equal to 0.83.

Dependent Variable: FARE3
Method: Least Squares
Date: 01/24/10 Time: 23:46
Sample (adjusted): 2 65
Included observations: 64 after adjustments
Convergence achieved after 18 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO3_16	-0.022901	0.010847	-2.111201	0.0389
C	457.4865	37.22225	12.29067	0.0000
AR(1)	0.820283	0.055311	14.83029	0.0000
MA(1)	0.416047	0.123755	3.361849	0.0014

R-squared	0.908452	Mean dependent var	408.5544
Adjusted R-squared	0.903874	S.D. dependent var	58.40192
S.E. of regression	18.10701	Akaike info criterion	8.690937
Sum squared resid	19671.82	Schwarz criterion	8.825867
Log likelihood	-274.1100	F-statistic	198.4642
Durbin-Watson stat	1.979693	Prob(F-statistic)	0.000000

Inverted AR Roots	.82
Inverted MA Roots	-.42

Figure 103: Delta Air Lines - regression on quarterly average fares. eco3_16 stands for the variable of Passenger Services.

- **US Airways:** The best model we find is a regression with the variable of the Transport Expenses, as for Continental and American Airlines (for the latter it is the Operational Expenses). We can check that the coefficient is negative as for the 2 other airlines, and the model explains almost 87% of the data on fares for the airline.

- **Southwest:** Finally, for the low-cost airline, we found as the best economic variable the Transport Revenue variable. Nevertheless, we can notice that here the sign for the economic coefficient is positive, which is contrary to the sign we expected for the other airlines (as we stressed that the variable of revenues and expenses are quite similar here). In fact we can try to explain this positive sign by the traditional strategy of a low-cost airline. For a traditional

Dependent Variable: FARE4
Method: Least Squares
Date: 01/25/10 Time: 00:12
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 21 iterations
Backcast: 2 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO4_4	0.002982	0.001077	2.769215	0.0078
C	548.7059	7.776075	70.56334	0.0000
@TREND	-3.300531	0.241603	-13.66098	0.0000
AR(1)	-0.544158	0.135673	-4.010819	0.0002
AR(2)	-0.452372	0.134231	-3.370111	0.0014
AR(3)	-0.424567	0.129552	-3.277199	0.0019
AR(4)	0.431459	0.128378	3.360859	0.0015
MA(1)	1.449148	0.081450	17.79187	0.0000
MA(2)	1.338471	0.110766	12.08383	0.0000
MA(3)	0.860623	0.066376	12.96582	0.0000

R-squared	0.966290	Mean dependent var	459.6851
Adjusted R-squared	0.960342	S.D. dependent var	55.37599
S.E. of regression	11.02780	Akaike info criterion	7.787536
Sum squared resid	6202.225	Schwarz criterion	8.133580
Log likelihood	-227.5198	F-statistic	162.4359
Durbin-Watson stat	1.963799	Prob(F-statistic)	0.000000

Inverted AR Roots	.49	-.02+.95i	-.02-.95i	-.98
Inverted MA Roots	-.23-.91i	-.23+.91i	-.98	

Figure 104: Northwest Airlines - regression on quarterly average fares. eco4_4 stands for the variable of Flying Operations.

airline, an increase in operational revenue strengthens the airline, which can then lower its fares and attract more passengers, maybe in order to catch a part of the passengers of the low-cost airlines. On the other hand, the well-known strategy of a low-cost airline is to charge the passenger for any additional service he wants to get. Thus we can assume that an increase in the operational expenses/revenues causes/is caused by an increase of the passenger fares (that is to say, an increase of the passenger service).

Dependent Variable: FARE5
Method: Least Squares
Date: 01/25/10 Time: 00:31
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 10 iterations
Backcast: 1 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO5_16	-0.070699	0.013389	-5.280222	0.0000
C	529.8724	13.79505	38.41031	0.0000
MA(1)	1.040072	0.108339	9.600208	0.0000
MA(2)	0.541555	0.107949	5.016770	0.0000

R-squared	0.833793	Mean dependent var	470.2444
Adjusted R-squared	0.825341	S.D. dependent var	52.41618
S.E. of regression	21.90586	Akaike info criterion	9.072772
Sum squared resid	28312.13	Schwarz criterion	9.208844
Log likelihood	-281.7923	F-statistic	98.65945
Durbin-Watson stat	1.932181	Prob(F-statistic)	0.000000
Inverted MA Roots	-.52+.52i	-.52-.52i	

Figure 105: United Airlines - regression on quarterly average fares. eco5_16 stands for the variable of Passenger Services.

Dependent Variable: FARE6
Method: Least Squares
Date: 01/25/10 Time: 00:27
Sample (adjusted): 1 65
Included observations: 65 after adjustments
Convergence achieved after 10 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO6_24	-0.010046	0.001929	-5.207162	0.0000
C	434.0422	9.626895	45.08642	0.0000
MA(1)	0.989819	0.068762	14.39478	0.0000
MA(2)	0.603541	0.051846	11.64096	0.0000

R-squared	0.867096	Mean dependent var	402.6931
Adjusted R-squared	0.860560	S.D. dependent var	53.94896
S.E. of regression	20.14545	Akaike info criterion	8.903398
Sum squared resid	24756.20	Schwarz criterion	9.037206
Log likelihood	-285.3604	F-statistic	132.6594
Durbin-Watson stat	2.190492	Prob(F-statistic)	0.000000
Inverted MA Roots	-.49+.60i	-.49-.60i	

Figure 106: US Airways - regression on quarterly average fares. eco6_24 stands for the variable of Transport Expenses.

Dependent Variable: FARE7
Method: Least Squares
Date: 01/25/10 Time: 00:28
Sample (adjusted): 23 65
Included observations: 43 after adjustments
Convergence achieved after 42 iterations
Backcast: 20 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ECO7_26	0.004920	0.000219	22.49275	0.0000
C	215.2520	1.037382	207.4954	0.0000
MA(1)	0.412661	0.093590	4.409263	0.0001
MA(2)	-0.543348	0.106619	-5.096145	0.0000
MA(3)	-0.732100	0.111419	-6.570674	0.0000

R-squared	0.851341	Mean dependent var	237.7842
Adjusted R-squared	0.835693	S.D. dependent var	14.63999
S.E. of regression	5.934298	Akaike info criterion	6.508319
Sum squared resid	1338.204	Schwarz criterion	6.713109
Log likelihood	-134.9288	F-statistic	54.40465
Durbin-Watson stat	1.994172	Prob(F-statistic)	0.000000

Inverted MA Roots	.96	-.68+.55i	-.68-.55i

Figure 107: Southwest Airlines - regression on quarterly average fares. eco7_26 stands for the variable of Transport Revenue.

6 Regressions for specific routes

It is interesting to see whether our "global" models for average fares and passengers for the entire domestic network in the U.S. could be used for specific round trips that are proposed by the 7 carriers examined in this project.

We can infer that on a specific route in the U.S. domestic market, the fares proposed by the airline would depend upon the competition between these airlines, the characterics of the route (distance, population data, airport data,...). In this section we want to assess how well a simple ARIMA model can work for a "given" route.

6.1 Chicago (ORD) - Atlanta (ATL)

The airports of Chicago O'Hare International (IATA airport code ORD) and Hartsfield-Jackson Atlanta International (IATA airport code ATL) were, in 2008, the 2 biggest airports in the U.S. in terms of enplanements. Indeed, we can assume that if we choose big airports, we will be able to get more general data, and not be stopped with specific characteristics we could have in smaller airports. We can also assume that the concentration of the market may be lower in these big airports.

We observe that the route between these 2 airports is only served by 3 carriers among the 7 airlines we study: American Airlines, Delta Air Lines, and United Airlines (see Figure 108). The other airlines may also propose some flights between Chicago O'Hare and Hatrsfield-Jackson Atlanta, but either they use code-sharing to do these flights, or they have not reported these flights to the DOT. We try to apply our best models (according to Section 4) to these average fares. We get the following regression outputs for the 3 carriers (Figures 109 to 111): we observe that these models do not seem to be the best regressions we can get to fit the fares between the 2 airports. The ARMA component of the model seems to be good, but the coefficients of trend or the dummy variable are insignificant for all regressions, according to the Student's t-statistics. As we said, this

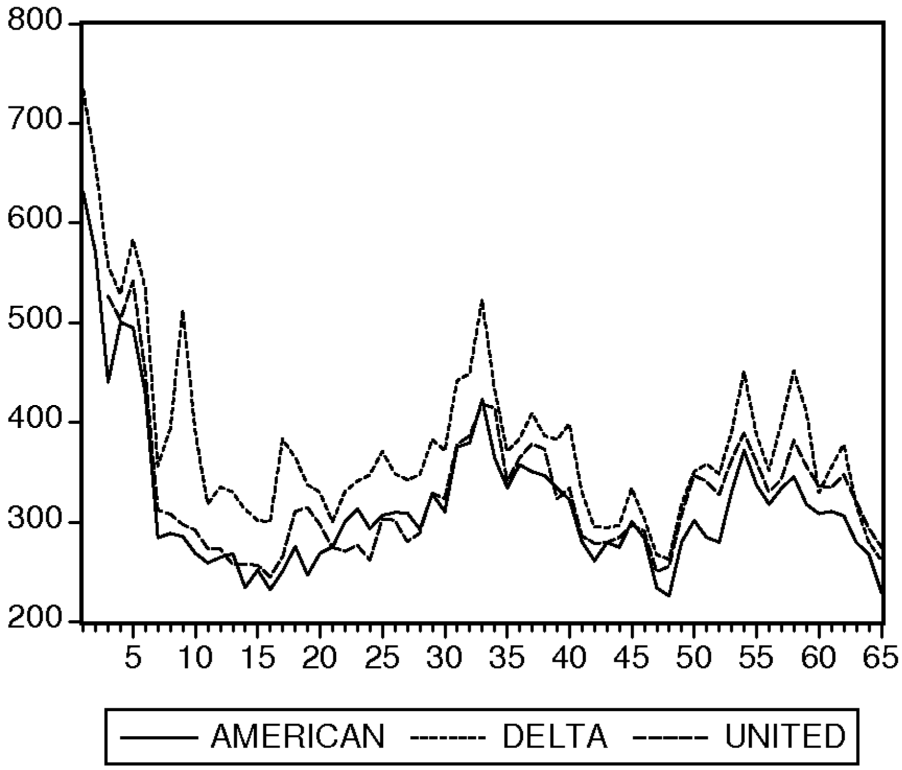


Figure 108: Quarterly average fares between Chicago (ORD) and Atlanta (ATL) ($ against time).

does not necessarily mean that our models are bad, but they seems to be too global to be able to capture the specificity of this particular route.

We may also assume that these global trends for the fares to decrease for these carriers, as well as the dummy variable modeling the impacts of September 11, 2001, may have been less important for the behaviour of the airlines between Chicago and Atlanta than between other U.S. cities such as New-York.

If we run the regressions for our best models for the number of passengers, we get the following results (Figures 112 to 114): curiously, as our models for these statistics on passengers appeared to be less accurate and precise than the best global regressions for average fares, we can see that these

Dependent Variable: AMERICAN
Method: Least Squares
Date: 01/22/10 Time: 19:14
Sample: 1 65
Included observations: 65
Convergence achieved after 14 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	344.6375	24.36217	14.14642	0.0000
DUMMY	33.72505	33.68619	1.001154	0.3208
@TREND	-1.538010	0.976840	-1.574476	0.1206
MA(1)	1.027928	0.067404	15.25027	0.0000
MA(2)	0.702527	0.057092	12.30515	0.0000

R-squared	0.827813	Mean dependent var	318.8659
Adjusted R-squared	0.816333	S.D. dependent var	76.34632
S.E. of regression	32.71923	Akaike info criterion	9.887607
Sum squared resid	64232.89	Schwarz criterion	10.05487
Log likelihood	-316.3472	F-statistic	72.11440
Durbin-Watson stat	1.670324	Prob(F-statistic)	0.000000
Inverted MA Roots	-.51-.66i	-.51+.66i	

Figure 109: American Airlines - regression on quarterly average fares.

models for American Airlines and for United Airlines are quite good. For the model for American Airlines, the model is stationary, all the coefficients are individually and jointly significant, the Durbin Watson statistic is quite low but there is no spike in the correlograms, the R^2 statistic is even higher than the coefficient of determination we have for the global regression (0.65 here against 0.62 for the regression on all passengers data).

For United Airlines, as we reject the dummy variable for the regression on fares, this coefficient is not very significant for our regression (but it remains significant at the 88% level of confidence, which is better than for the model for the average fares). All the other coefficients are individually significant at the 95% level of confidence.

For Delta Air Lines, our model does not seem to be the best one for the statistic on passengers for the airline. In fact, we can notice that in our regression on the number of passengers for Delta Air Lines, we did not consider any seasonality effect for the airline in our best model according

Dependent Variable: DELTA
Method: Least Squares
Date: 01/22/10 Time: 19:16
Sample (adjusted): 2 65
Included observations: 64 after adjustments
Convergence achieved after 13 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	409.8577	40.40135	10.14465	0.0000
@TREND	-1.302440	1.032429	-1.261530	0.2120
AR(1)	0.557524	0.110048	5.066178	0.0000
MA(1)	0.495031	0.130422	3.795600	0.0003

R-squared	0.698170	Mean dependent var	375.3657
Adjusted R-squared	0.683078	S.D. dependent var	79.70510
S.E. of regression	44.87062	Akaike info criterion	10.50591
Sum squared resid	120802.3	Schwarz criterion	10.64084
Log likelihood	-332.1890	F-statistic	46.26243
Durbin-Watson stat	2.212073	Prob(F-statistic)	0.000000

Inverted AR Roots	.56
Inverted MA Roots	-.50

Figure 110: Delta Air Lines - regression on quarterly average fares.

to the SIC. Here we can see that the coefficients AR(1) and AR(2) do not seem to be significant, but this result could also mean that the data would better fit a seasonal autoregressive component rather than individual AR variables. In fact, if we try to find the best model for the number of passengers of Delta between these airports, we find with EViews6 an ARMA(2,1), with a seasonal variable SAR(4) and a trend variable (see Figure 115). This result seems to be quite a good model, except for the invertibility of the regression. All coefficients are significant, the Durbin-Watson statistic is higher than 2, there is no sign of autocorrelation in the correlograms, the model is stationary, and the R^2 statistic is very high (0.90). As a consequence, we can confirm that the seasonality of the number of passengers really matters in the models for the number of passengers.

Dependent Variable: UNITED
Method: Least Squares
Date: 01/22/10 Time: 19:20
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 17 iterations
Backcast: 1 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	316.6032	14.44881	21.91207	0.0000
DUMMY	11.70167	19.96145	0.586214	0.5600
MA(1)	1.075850	0.060075	17.90849	0.0000
MA(2)	0.443546	0.077294	5.738405	0.0000

R-squared	0.741212	Mean dependent var	326.8441
Adjusted R-squared	0.728054	S.D. dependent var	63.47751
S.E. of regression	33.10256	Akaike info criterion	9.898485
Sum squared resid	64650.98	Schwarz criterion	10.03456
Log likelihood	-307.8023	F-statistic	56.32870
Durbin-Watson stat	1.695740	Prob(F-statistic)	0.000000
Inverted MA Roots	-.54-.39i	-.54+.39i	

Figure 111: United Airlines - regression on quarterly average fares.

Dependent Variable: PAX1_R
Method: Least Squares
Date: 01/22/10 Time: 20:45
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 11 iterations
Backcast: 3 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1625.691	186.7098	8.707048	0.0000
AR(4)	0.494290	0.134612	3.671966	0.0005
MA(1)	0.684753	0.124421	5.503531	0.0000
MA(2)	0.468036	0.128473	3.643073	0.0006

R-squared	0.652094	Mean dependent var	1633.131
Adjusted R-squared	0.633783	S.D. dependent var	565.4451
S.E. of regression	342.1839	Akaike info criterion	14.57190
Sum squared resid	6674119.	Schwarz criterion	14.71032
Log likelihood	-440.4429	F-statistic	35.61247
Durbin-Watson stat	1.778421	Prob(F-statistic)	0.000000

Inverted AR Roots	.84	.00-.84i	-.00+.84i	-.84
Inverted MA Roots	-.34-.59i	-.34+.59i		

Figure 112: American Airlines - regression on the number of passengers per quarter.

Dependent Variable: PAX3_R
Method: Least Squares
Date: 01/22/10 Time: 20:50
Sample (adjusted): 4 65
Included observations: 62 after adjustments
Convergence achieved after 9 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4922.134	665.2754	7.398640	0.0000
DUMMY	-1984.482	691.7610	-2.868739	0.0058
AR(1)	0.428583	0.136457	3.140799	0.0027
AR(2)	0.153651	0.146554	1.048425	0.2989
AR(3)	0.216699	0.133430	1.624071	0.1099

R-squared	0.695189	Mean dependent var	3897.839
Adjusted R-squared	0.673799	S.D. dependent var	1413.615
S.E. of regression	807.3718	Akaike info criterion	16.30265
Sum squared resid	37155410	Schwarz criterion	16.47420
Log likelihood	-500.3822	F-statistic	32.50033
Durbin-Watson stat	2.183800	Prob(F-statistic)	0.000000

Inverted AR Roots	.88	-.23+.44i	-.23-.44i

Figure 113: Delta Air Lines - regression on the number of passengers per quarter.

Dependent Variable: PAX5_R
Method: Least Squares
Date: 01/22/10 Time: 20:49
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 23 iterations
Backcast: 5 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3783.720	315.7964	11.98152	0.0000
DUMMY	-515.8166	319.3217	-1.615351	0.1122
@TREND	-33.90916	11.05596	-3.067047	0.0034
AR(4)	0.281860	0.100272	2.810952	0.0069
MA(1)	0.271190	0.135097	2.007367	0.0498
MA(2)	0.296377	0.134230	2.207978	0.0316

R-squared	0.841305	Mean dependent var	2287.169
Adjusted R-squared	0.826334	S.D. dependent var	856.7830
S.E. of regression	357.0496	Akaike info criterion	14.68977
Sum squared resid	6756676.	Schwarz criterion	14.90105
Log likelihood	-427.3482	F-statistic	56.19477
Durbin-Watson stat	1.849356	Prob(F-statistic)	0.000000

Inverted AR Roots	.73	.00-.73i
Inverted MA Roots	-.14-.53i	-.14+.53i

Figure 114: United Airlines - regression on the number of passengers per quarter.

Dependent Variable: PAX3_R
Method: Least Squares
Date: 01/22/10 Time: 21:15
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 67 iterations
Backcast: OFF (Roots of MA process too large)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9766.443	3413.447	2.861168	0.0060
@TREND	-129.5204	61.88975	-2.092760	0.0412
AR(1)	1.659360	0.110624	15.00001	0.0000
AR(2)	-0.721732	0.106080	-6.803671	0.0000
SAR(4)	0.754986	0.085112	8.870495	0.0000
MA(1)	-1.192886	0.151067	-7.896419	0.0000

R-squared	0.903605	Mean dependent var	3937.305
Adjusted R-squared	0.894511	S.D. dependent var	1436.651
S.E. of regression	466.6113	Akaike info criterion	15.22501
Sum squared resid	11539481	Schwarz criterion	15.43629
Log likelihood	-443.1379	F-statistic	99.36384
Durbin-Watson stat	2.202642	Prob(F-statistic)	0.000000

Inverted AR Roots	.93	.83+.18i	.83-.18i	.00-.93i
	.00+.93i	-.93		
Inverted MA Roots	1.19			
	Estimated MA process is noninvertible			

Figure 115: Delta Air Lines - regression on the number of passengers per quarter.

6.2 Los Angeles (LAX) - New-York (JFK)

We now consider 2 other big airports, Los Angeles International (IATA airpot code LAX) and John F Kennedy International (IATA airport code JFK), in Los Angeles and New-York respectively. The same airlines as for the previous route are competing between these 2 airports (see Figure 116). If

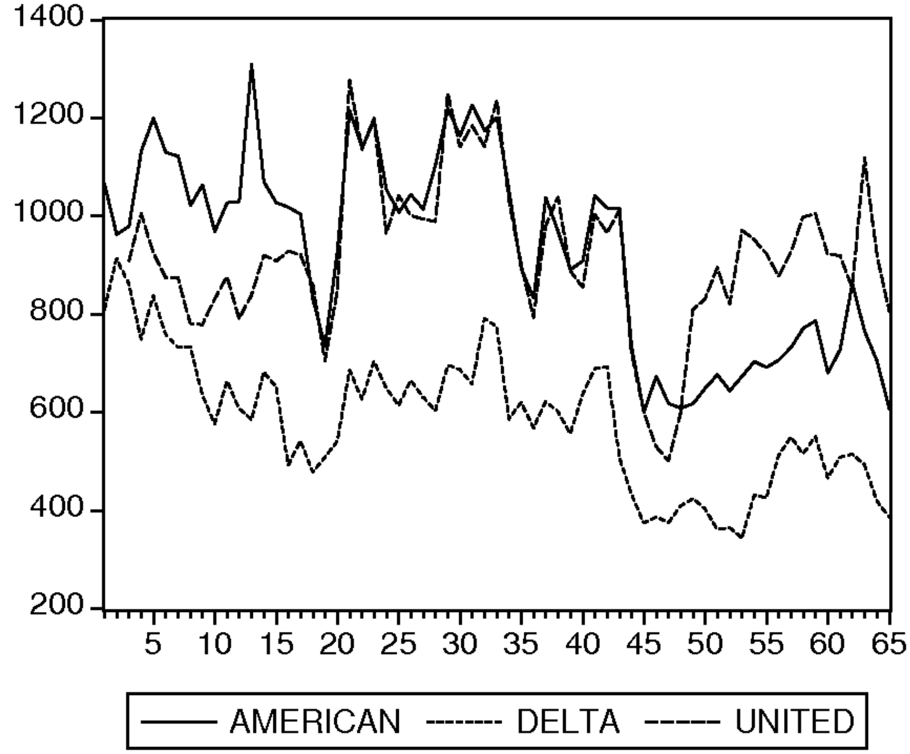


Figure 116: Quarterly average fares between Los Angeles (LAX) and New-York (JFK) ($ against time).

we run the same regressions as for the previous route for the fares of the 3 airlines (Figure 117 to 119), we can then observe that the conclusions are not the same as for the previous route: we can observe that all the coefficients are significant for the regression of United. For Delta Airlines, the trend variable also seems to be significant, as the dummy variable is for American Airlines. The model for United even seems to be a good one, but the low R^2 statistic may be a sign that this

ARMA model does not explain all the variations of the data.

Dependent Variable: FARE1
Method: Least Squares
Date: 01/22/10 Time: 22:05
Sample: 1 65
Included observations: 65
Convergence achieved after 11 iterations
Backcast: -1 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1111.544	58.38891	19.03691	0.0000
@TREND	-2.668128	2.547093	-1.047519	0.2991
DUMMY	-216.1821	92.76353	-2.330465	0.0232
MA(1)	0.767416	0.116795	6.570635	0.0000
MA(2)	0.423425	0.116223	3.643226	0.0006

R-squared	0.779936	Mean dependent var	927.1746
Adjusted R-squared	0.765265	S.D. dependent var	198.4905
S.E. of regression	96.16759	Akaike info criterion	12.04387
Sum squared resid	554892.3	Schwarz criterion	12.21113
Log likelihood	-386.4256	F-statistic	53.16198
Durbin-Watson stat	1.927842	Prob(F-statistic)	0.000000

Inverted MA Roots	-.38+.53i	-.38-.53i

Figure 117: American Airlines - regression on quarterly average fares.

For the regressions on the number of passengers (Figure 120 to 122), the results are also different from the previous conclusions. We get quite a good model for American Airlines, as all the coefficients of the regression are significant. There is no problem of stationarity or autocorrelation, and the coefficient of determination R^2 is also higher than for the global regression on passengers (0.70 here). For Delta Airlines, again we may infer that the seasonality of the number of passengers is the cause of the rejection of the variable AR(2) and AR(3), and the model is consequently not as good as the global regression. On the contrary, for United Airlines, the model seems to reject the seasonality of the number of passengers, whereas we kept this variable for the route between Chicago and Atlanta. Again it stresses the importance of taking into account local effects and the characteristics of the routes in order to better fit the data we have from our databases. Here we notice that the dummy variable seems to be important in the regression, whereas this coefficient

Dependent Variable: FARE3
Method: Least Squares
Date: 01/22/10 Time: 22:06
Sample (adjusted): 2 65
Included observations: 64 after adjustments
Convergence achieved after 8 iterations
Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	732.4027	70.37203	10.40758	0.0000
@TREND	-4.901531	1.745794	-2.807623	0.0067
AR(1)	0.760075	0.113175	6.715953	0.0000
MA(1)	-0.077398	0.175711	-0.440488	0.6612

R-squared	0.767188	Mean dependent var	579.0942
Adjusted R-squared	0.755548	S.D. dependent var	134.5663
S.E. of regression	66.53241	Akaike info criterion	11.29372
Sum squared resid	265593.7	Schwarz criterion	11.42865
Log likelihood	-357.3989	F-statistic	65.90636
Durbin-Watson stat	1.914511	Prob(F-statistic)	0.000000

Inverted AR Roots	.76
Inverted MA Roots	.08

Figure 118: Delta Air Lines - regression on quarterly average fares.

was not significant on the first route studied.

Dependent Variable: FARE5
Method: Least Squares
Date: 01/22/10 Time: 22:04
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 11 iterations
Backcast: 1 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	979.1359	43.94639	22.28023	0.0000
DUMMY	-123.4689	61.24552	-2.015966	0.0484
MA(1)	0.737380	0.106797	6.904489	0.0000
MA(2)	0.526463	0.114141	4.612393	0.0000

R-squared	0.522884	Mean dependent var	922.2468
Adjusted R-squared	0.498624	S.D. dependent var	158.1343
S.E. of regression	111.9716	Akaike info criterion	12.33575
Sum squared resid	739721.0	Schwarz criterion	12.47183
Log likelihood	-384.5763	F-statistic	21.55320
Durbin-Watson stat	1.941271	Prob(F-statistic)	0.000000
Inverted MA Roots	-.37+.62i	-.37-.62i	

Figure 119: United Airlines - regression on quarterly average fares.

Dependent Variable: FARE1
Method: Least Squares
Date: 01/22/10 Time: 22:22
Sample (adjusted): 5 65
Included observations: 61 after adjustments
Convergence achieved after 69 iterations
Backcast: 3 4

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	900.0526	52.20771	17.23984	0.0000
AR(4)	0.383516	0.137257	2.794135	0.0071
MA(1)	0.693463	0.124630	5.564155	0.0000
MA(2)	0.494549	0.118599	4.169912	0.0001

R-squared	0.699325	Mean dependent var	920.1204
Adjusted R-squared	0.683500	S.D. dependent var	202.2118
S.E. of regression	113.7610	Akaike info criterion	12.36940
Sum squared resid	737669.6	Schwarz criterion	12.50782
Log likelihood	-373.2668	F-statistic	44.19113
Durbin-Watson stat	1.601820	Prob(F-statistic)	0.000000

Inverted AR Roots	.79	-.00+.79i	-.00-.79i	-.79
Inverted MA Roots	-.35-.61i	-.35+.61i		

Figure 120: American Airlines - regression on the number of passengers per quarter.

Dependent Variable: FARE3
Method: Least Squares
Date: 01/22/10 Time: 22:23
Sample (adjusted): 4 65
Included observations: 62 after adjustments
Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	611.6708	45.20144	13.53211	0.0000
DUMMY	-123.8819	53.73269	-2.305521	0.0248
AR(1)	0.695335	0.132409	5.251425	0.0000
AR(2)	-0.052021	0.160163	-0.324799	0.7465
AR(3)	0.092662	0.126210	0.734190	0.4658

R-squared	0.737159	Mean dependent var	569.1426
Adjusted R-squared	0.718714	S.D. dependent var	124.3354
S.E. of regression	65.94297	Akaike info criterion	11.29266
Sum squared resid	247863.1	Schwarz criterion	11.46421
Log likelihood	-345.0726	F-statistic	39.96534
Durbin-Watson stat	1.941385	Prob(F-statistic)	0.000000

Inverted AR Roots	.78	-.04+.34i	-.04-.34i

Figure 121: Delta Air Lines - regression on the number of passengers per quarter.

Dependent Variable: FARE5
Method: Least Squares
Date: 01/22/10 Time: 22:26
Sample (adjusted): 7 65
Included observations: 59 after adjustments
Convergence achieved after 11 iterations
Backcast: 5 6

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	851.7924	92.43592	9.214950	0.0000
DUMMY	-293.7036	115.3611	-2.545951	0.0138
@TREND	6.147600	3.574554	1.719823	0.0913
AR(4)	0.057623	0.151746	0.379733	0.7057
MA(1)	0.713245	0.128705	5.541704	0.0000
MA(2)	0.511023	0.130599	3.912911	0.0003

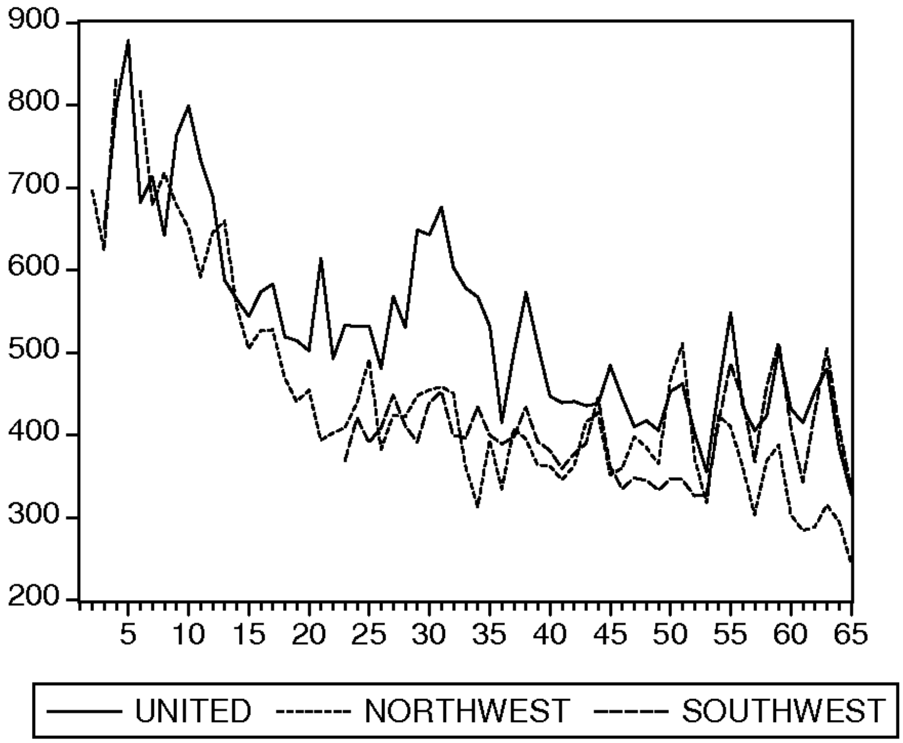
R-squared	0.562606	Mean dependent var	921.8376
Adjusted R-squared	0.521343	S.D. dependent var	163.0129
S.E. of regression	112.7806	Akaike info criterion	12.38491
Sum squared resid	674131.4	Schwarz criterion	12.59618
Log likelihood	-359.3548	F-statistic	13.63447
Durbin-Watson stat	1.975700	Prob(F-statistic)	0.000000

Inverted AR Roots	.49	-.00+.49i	-.00-.49i	-.49
Inverted MA Roots	-.36+.62i	-.36-.62i		

Figure 122: United Airlines - regression on the number of passengers per quarter.

6.3 Baltimore (BWI) - Seattle (SEA)

What about the low-cost airline Southwest? Its specificity may also lead to disruption of our models for fares and passengers in case of competition with the major airlines. Consider for example the route between Baltimore/Washington International Airport (BWI) and Seattle-Tacoma International Airport. This route has been offered by the 2 major carriers Northwest, United since 1993, and also by Southwest If we consider the fares for this route (Figure 123), we can observe that there seems to be a trend for the fares to continuously decrease after 1993 probably to align with the fares proposed by the low-cost airline. In addition we can also observe the trend for the fares of Northwest to keep on decreasing below the fares of Southwest, following the global decreasing trend of fares we observed in the previous part for Northwest. If we try to run the regressions on the



,

Figure 123: Quarterly average fares between Baltimore (BWI) and Seattle (SEA) ($ against time).

fares for these 3 airlines, we get the following results (Figures 124 to 126): we can observe similar results as for the previous routes studied: we can see that the coefficient of trend is significant for the regression on the fares of Northwest Airlines, as is the the dummy variable for United Airlines.

Dependent Variable: FARE4
Method: Least Squares
Date: 02/03/10 Time: 18:30
Sample (adjusted): 10 65
Included observations: 56 after adjustments
Convergence achieved after 33 iterations
Backcast: 7 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	467.8731	45.96929	10.17795	0.0000
@TREND	-2.125384	0.919522	-2.311401	0.0252
AR(1)	1.246599	0.140570	8.868170	0.0000
AR(2)	-1.239577	0.109543	-11.31585	0.0000
AR(3)	1.282528	0.132087	9.709718	0.0000
AR(4)	-0.434433	0.121779	-3.567379	0.0008
MA(1)	-0.730179	0.055077	-13.25737	0.0000
MA(2)	0.768951	0.053460	14.38374	0.0000
MA(3)	-0.959601	0.038778	-24.74629	0.0000

R-squared	0.845225	Mean dependent var	414.4536
Adjusted R-squared	0.818880	S.D. dependent var	90.97878
S.E. of regression	38.71890	Akaike info criterion	10.29676
Sum squared resid	70460.22	Schwarz criterion	10.62226
Log likelihood	-279.3092	F-statistic	32.08331
Durbin-Watson stat	2.004095	Prob(F-statistic)	0.000000

Inverted AR Roots	.86	.52	-.07+.98i	-.07-.98i
Inverted MA Roots	.96	-.12+.99i	-.12-.99i	

Figure 124: Northwest Airlines - regression on quarterly average fares.

These 2 regressions remain stationary and do not show problems of autocorrelation in the residuals, but we may assume that again an ARIMA model is not sufficiently able to represent the behaviour of passenger fares on a route.

For Southwest Airlines, the model reveals the increasing trend for the fares, and a dummy variable after the terorist events of September 2001 (however this possible impact is quite low in

Dependent Variable: FARE5
Method: Least Squares
Date: 02/03/10 Time: 18:32
Sample (adjusted): 3 65
Included observations: 63 after adjustments
Convergence achieved after 11 iterations
Backcast: 1 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	620.4826	24.44770	25.38000	0.0000
DUMMY	-183.6631	33.87270	-5.422156	0.0000
MA(1)	0.828315	0.090572	9.145405	0.0000
MA(2)	0.555726	0.093072	5.970957	0.0000

R-squared	0.756772	Mean dependent var	532.9023
Adjusted R-squared	0.744405	S.D. dependent var	117.0788
S.E. of regression	59.19086	Akaike info criterion	11.06080
Sum squared resid	206709.9	Schwarz criterion	11.19687
Log likelihood	-344.4151	F-statistic	61.19030
Durbin-Watson stat	1.929481	Prob(F-statistic)	0.000000
Inverted MA Roots	-.41-.62i	-.41+.62i	

Figure 125: United Airlines - regression on quarterly average fares.

comparison with the global impact we observed for the major airlines). Nevertheless, it appears difficult to find a quite good model with all the coefficients remaining significant.

The same conclusions can be drawn for the statistic on the number of passengers: we can observe that the dummy variable and trend variable appear to be insignificant for most airlines, and the R^2 statistics are very low for all the regressions.

Dependent Variable: FARE7
Method: Least Squares
Date: 02/03/10 Time: 18:32
Sample (adjusted): 23 65
Included observations: 43 after adjustments
Convergence achieved after 22 iterations
Backcast: 20 22

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	385.8249	36.67719	10.51948	0.0000
@TREND	0.795493	1.160668	0.685375	0.4974
DUMMY	-31.69342	31.10817	-1.018814	0.3149
MA(1)	0.914172	0.159901	5.717103	0.0000
MA(2)	0.049678	0.216083	0.229902	0.8194
MA(3)	-0.354064	0.145176	-2.438863	0.0197

R-squared	0.507488	Mean dependent var	398.2017
Adjusted R-squared	0.440933	S.D. dependent var	47.21875
S.E. of regression	35.30584	Akaike info criterion	10.09476
Sum squared resid	46120.58	Schwarz criterion	10.34051
Log likelihood	-211.0374	F-statistic	7.625026
Durbin-Watson stat	2.039542	Prob(F-statistic)	0.000052
Inverted MA Roots	.49	-.70+.49i	-.70-.49i

Figure 126: Southwest Airlines - regression on quarterly average fares.

7 Purely economic models

We saw in the previous sections some time series models for quarterly average fares for our 7 airlines. However, we can suppose that the fares proposed by an airline on a given route also depend on the characteristics of the market, and are not purely caused by the economic state of the airline. The fares may also be a function of the length of the route, the kind of route (e.g. route for leisure or business travellers), the population living at the endpoints of the route, and so on. The IENAC06TE study led by Jonathan Cobb and Nico Metzger, "Entry Strategy for Southwest Airlines", tries to find the models which could explain the fares of the low-cost airline for a specific route.[19]

We would like to observe whether it is possible to add some of the variables used in that study on Southwest to some basic panel data, taking into account the possible trend on the fares and the impacts of September 11, 2001 (with our dummy variable). Then we have to estimate, for a given carrier, the following equation:

$$\begin{aligned}\mathrm{Fare}(t) = \alpha_0 + \alpha_1 \times t + \alpha_2 \times \mathrm{D}(t) + \alpha_3 \times \text{distance} + \alpha_4 \times \text{distance} + \alpha_5 \times \text{population} \\ + \alpha_6 \times \text{leisure} + \alpha_7 \times \text{business} + \alpha_8 \times \mathrm{GDP} + \varepsilon(t),\end{aligned} \tag{29}$$

where "distance", "population" and "GDP" are discrete variables chosen by Jonathan Cobb and Nico Metzger to model the characteristics of the market around the endpoints of the route. $\mathrm{D}(t)$ is a dummy variable used to model the U.S. terrorist attacks in 2001 (equal to 1 at and after period 35, and 0 otherwise), and "leisure" and "business" are also dummy variables, taking the value 1 if the route can be considered (according to the criteria developed in the study of Cobb and Metzger) as a "leisure" or "business" route. Here, we also include the variable of time in the regression to capture possible trends.

[19] Research Project "Entry Strategy for Southwest Airlines", by Jonathan Cobb and Nico Metzger, supervised by Steve Lawford [6].

We get the following regression output from EViews6: we can observe that in fact these models are quite good! Most of coefficients of the 7 regressions are individually and jointly significant at the 95% and very often at the 99% levels, except for the population variable for Southwest and Northwest, and the Business or Leisure variable for Continental Airlines and Northwest.

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:00
Sample: 1 31960
Included observations: 31960

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	379.5481	11.81848	32.11481	0.0000
T	-2.290994	0.083773	-27.34761	0.0000
DISTANCE	0.084423	0.000661	127.6961	0.0000
POPN/1000	-0.029912	0.001619	-18.47698	0.0000
LEISURE	-75.06033	2.213105	-33.91630	0.0000
BUSINESS	24.31567	11.68333	2.081228	0.0374
GDP/1000	0.696193	0.033023	21.08175	0.0000
SEP11	-87.25337	3.139054	-27.79608	0.0000

R-squared	0.472109	Mean dependent var	488.1838
Adjusted R-squared	0.471994	S.D. dependent var	191.3783
S.E. of regression	139.0632	Akaike info criterion	12.70798
Sum squared resid	6.18E+08	Schwarz criterion	12.71008
Log likelihood	-203065.6	F-statistic	4082.239
Durbin-Watson stat	0.310945	Prob(F-statistic)	0.000000

Figure 127: American Airlines - economic regression on quarterly average fares.

We note that if we do not include these variables in the regression, we only get significant variables in all our models. The R^2 statistics of the models are quite high. Indeed, the R^2 statistic for American Airlines is equal to 0.47, meaning that these variables already explain 47% of the fare setting behaviour for this carrier! This statistic is equal to 0.31 for Continental, 0.36 for Delta, 0.17 for Northwest, 0.54 for United,0.22 for US Airways and 0.52 for Southwest Airlines.

Again, we can observe a clear difference between the trend variable for the major carriers and Southwest: whereas the coefficient for the trend variable is lower than -1.70 dollars per quarter for all the major carriers but Continental, this coefficient is equal to 0.81 for Southwest. This trend

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:01
Sample: 1 15184
Included observations: 15184

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	309.3503	18.44364	16.77274	0.0000
T	-0.303479	0.109420	-2.773528	0.0056
DISTANCE	0.063429	0.000934	67.91425	0.0000
POPN/1000	-0.045714	0.002243	-20.38514	0.0000
LEISURE	-53.28485	2.691832	-19.79501	0.0000
BUSINESS	29.28988	18.37931	1.593633	0.1110
GDP/1000	1.051212	0.044194	23.78629	0.0000
SEP11	-58.11834	4.106555	-14.15258	0.0000

R-squared	0.314461	Mean dependent var	444.4828
Adjusted R-squared	0.314145	S.D. dependent var	154.4615
S.E. of regression	127.9194	Akaike info criterion	12.54120
Sum squared resid	2.48E+08	Schwarz criterion	12.54522
Log likelihood	-95204.83	F-statistic	994.4739
Durbin-Watson stat	0.366560	Prob(F-statistic)	0.000000

Figure 128: Continental Airlines - economic regression on quarterly average fares.

shows the competition between these 2 models of airlines, and we may suggest according to these statistics, that there is a trend for the difference of fare setting between a low-cost airline and a major one, to be reduced.

The coefficient of the route length is positive for all the airlines: indeed, we can assume that the fares will be higher on a longer route.

We can observe that the coefficient for the population variable is negative, meaning that the fares tend to be higher between less populated areas. Indeed, it seems logical that in more populated areas, where demand and competition between airlines is higher, the fares will decrease. However we can observe that we reject this coefficient for Southwest Airlines, which can also be interpreted as a characteristic for the low-cost airline, to use secondary airports, and point-to-point services rather than hub-and-spoke networks. Using these secondary airports, Southwest also serves less populated areas in the neighbourhood of big cities where major airlines develop their hub airports.

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:03
Sample: 1 40090
Included observations: 40090

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	435.6770	6.049651	72.01688	0.0000
T	-3.794347	0.071961	-52.72758	0.0000
DISTANCE	0.071834	0.000605	118.7003	0.0000
POPN/1000	-0.042562	0.001961	-21.70131	0.0000
LEISURE	-67.63937	1.597287	-42.34642	0.0000
BUSINESS	27.85667	5.816673	4.789107	0.0000
GDP/1000	0.798462	0.040062	19.93091	0.0000
SEP11	-14.17621	2.495536	-5.680627	0.0000

R-squared	0.360563	Mean dependent var	454.4555
Adjusted R-squared	0.360452	S.D. dependent var	166.4448
S.E. of regression	133.1089	Akaike info criterion	12.62041
Sum squared resid	7.10E+08	Schwarz criterion	12.62213
Log likelihood	-252968.1	F-statistic	3228.757
Durbin-Watson stat	0.320995	Prob(F-statistic)	0.000000

Figure 129: Delta Air Lines - economic regression on quarterly average fares.

The “leisure” and “business” variables show the trend for major airlines to decrease their fares on specific leisure routes, and increase these fares on business routes. For Southwest again, the traditional strategy for the low-cost airline is to not discriminate between business or leisure travellers but offer them the lower fare for their trip. Consequently, we can observe that the low-cost airline is the only airline with both coefficients negative for the variables leisure and business.

Again the low-cost airline is traditionaly not supposed to take into account the GDP on a given region to fix a higher or lower fare than for another U.S. region. Thus we can see that the coefficient for GDP is almost equal to 0 for Southwest. On the contrary, as this coefficient is definitely positive for the major airlines, we can assume that they seem to use this statistic in their fare setting to propose enforce higher fares in richer regions of the USA!

The impact of the terrorist attacks in September 11 may be one of the only coefficients with a similar impact on the low-cost airline and the legacy carriers. These events seem to have forced the

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:04
Sample: 1 14264
Included observations: 14264

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	434.4327	14.90564	29.14553	0.0000
T	-2.252883	0.122767	-18.35084	0.0000
DISTANCE	0.023312	0.000952	24.47742	0.0000
POPN/1000	-0.004622	0.003287	-1.406055	0.1597
LEISURE	1.154329	2.621522	0.440328	0.6597
BUSINESS	52.21146	14.43909	3.615979	0.0003
GDP/1000	0.064672	0.067278	0.961259	0.3364
SEP11	-32.22625	4.437407	-7.262405	0.0000

R-squared	0.167501	Mean dependent var	447.7542
Adjusted R-squared	0.167092	S.D. dependent var	147.8276
S.E. of regression	134.9131	Akaike info criterion	12.64770
Sum squared resid	2.59E+08	Schwarz criterion	12.65194
Log likelihood	-90195.39	F-statistic	409.7629
Durbin-Watson stat	0.290115	Prob(F-statistic)	0.000000

Figure 130: Northwest Airlines - economic regression on quarterly average fares.

airlines to adapt their fares to a decreasing demand. This decrease of the fares appears to be very high for American Airlines and United Airlines (minus $87 and $89 respectively for the average fares of the 2 airlines), whereas Delta and Southwest seem to have been able to lower the impact of the terrorist attacks on their fare strategy (which can also be explained by the fact that Southwest in particular already had the lowest fares among these airlines before September 2001).

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:04
Sample: 1 24728
Included observations: 24728

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	270.5218	15.04574	17.97996	0.0000
T	-1.700538	0.098881	-17.19781	0.0000
DISTANCE	0.099579	0.000691	144.0634	0.0000
POPN/1000	-0.055815	0.001845	-30.24770	0.0000
LEISURE	-89.06896	2.601667	-34.23535	0.0000
BUSINESS	74.95982	14.93683	5.018456	0.0000
GDP/1000	1.266880	0.037283	33.98007	0.0000
SEP11	-89.75342	3.440959	-26.08384	0.0000

R-squared	0.545722	Mean dependent var	508.3681
Adjusted R-squared	0.545593	S.D. dependent var	210.2143
S.E. of regression	141.7049	Akaike info criterion	12.74569
Sum squared resid	4.96E+08	Schwarz criterion	12.74832
Log likelihood	-157579.8	F-statistic	4242.283
Durbin-Watson stat	0.333079	Prob(F-statistic)	0.000000

Figure 131: United Airlines - economic regression on quarterly average fares.

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:05
Sample: 1 27230
Included observations: 27230

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	469.9082	6.293602	74.66443	0.0000
T	-1.971294	0.083367	-23.64598	0.0000
DISTANCE	0.033739	0.000859	39.26543	0.0000
POPN/1000	-0.083990	0.002229	-37.68008	0.0000
LEISURE	-90.25694	2.052832	-43.96704	0.0000
BUSINESS	20.49648	6.071357	3.375931	0.0007
GDP/1000	1.569493	0.044537	35.23984	0.0000
SEP11	-39.05656	3.014560	-12.95597	0.0000

R-squared	0.217890	Mean dependent var	424.9593
Adjusted R-squared	0.217689	S.D. dependent var	146.0084
S.E. of regression	129.1420	Akaike info criterion	12.56000
Sum squared resid	4.54E+08	Schwarz criterion	12.56241
Log likelihood	-170996.3	F-statistic	1083.409
Durbin-Watson stat	0.318775	Prob(F-statistic)	0.000000

Figure 132: US Airways - economic regression on quarterly average fares.

Dependent Variable: FARE
Method: Least Squares
Date: 01/24/10 Time: 19:06
Sample: 1 30324
Included observations: 30324

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	214.7214	2.594307	82.76636	0.0000
T	0.809883	0.040662	19.91738	0.0000
DISTANCE	0.060413	0.000333	181.4040	0.0000
POPN/1000	-0.000445	0.000916	-0.485621	0.6272
LEISURE	-22.89062	0.833207	-27.47291	0.0000
BUSINESS	-16.20416	2.238844	-7.237737	0.0000
GDP/1000	-0.078174	0.018652	-4.191260	0.0000
SEP11	-20.48762	1.180283	-17.35823	0.0000

R-squared	0.529839	Mean dependent var	302.9831
Adjusted R-squared	0.529730	S.D. dependent var	82.57083
S.E. of regression	56.62396	Akaike info criterion	10.91101
Sum squared resid	97201369	Schwarz criterion	10.91320
Log likelihood	-165424.7	F-statistic	4880.570
Durbin-Watson stat	0.335705	Prob(F-statistic)	0.000000

Figure 133: Southwest Airlines - economic regression on quarterly average fares.

8 Concluding remarks

The objective of this project was to conduct research into the modelling of the dynamic strategic behaviour of U.S. major carriers since 1993. The first step consisted in filtering the data that we were able to get from the U.S. Department of Transportation on the coupons and tickets bought by U.S. travellers. In particular, we decided to use the quarterly average fares proposed by the airlines for all round trips inside the U.S. domestic market, and the quarterly number of passengers which travelled on these routes since 1993. The data treatment and descriptive analysis formed an important first step of the work.

Then we supposed that the fares set by the airlines may depend on the past evolution of fares, and on the impact of economical or political events such as the U.S. terrorist attack in 2001. We built software to select and test rigorous time series models, we observed that we were able to capture a large amount of the fare setting strategies for these airlines. In particular, these models enabled us to clarify and quantify the different strategies between legacy airlines and a low-cost carrier such as Southwest Airlines. We also saw the impact that the September 11, 2001, had on each of the carriers, and the possible trends for these fares to decrease (especially for Northwest Airlines), or to increase as they seem to do for Southwest. These time series models also enabled us to make some robust static and dynamic forecasts on a short-term horizon, as to how fares may behave in the next 1 to 2 years, and how the number of passengers may evolve. This analysis has been fully automated using customized code, which formed a significant part of the modelling work, before we could begin interpretation of the results. It would also be a vital part of a repeated model estimation procedure, such as is often found in the private sector and other business settings.

Then we attempted to see whether the economic state of the carriers could be correlated with their fare setting behaviour. Using our best time series models, we included financial variables such as operational revenues or expenses, the cost of passenger services or the total number of employees in a given airline, and saw that these variables can in fact explain most of the trends and drops observed in the purely statistical models.

Finally, we wished to see whether purely economic models could be used to explain the fare setting. Using variables on route characteristics such as distance, whether the route is used more for business or leisure purposes, and the characteristics of the population and the GDP at the endpoints of this route, we saw that we could explain 30 to 40% of what is happening on these routes!

However, none of these models can be perfect. They could be improved in a number of ways. For example, it would be interesting to model more carefully the competition between the major airlines, to capture more realistically the strategy of the airlines in the U.S. market. It could be very revealing to observe what competitors do on a given route and to include the entry decisions of competitors. One of the difficulties we noticed with the major airlines is that there are also many existing and nascent code-sharing agreements between these carriers and smaller ones, which allow the carriers to extend their networks without necessarily being involved in the fare setting, nor in the strategy of these smaller carriers. These code-sharing effects are very difficult to capture as these agreements are temporary, but they may also have an impact on the behaviour of airlines and on their competition with other airlines.

8.1 Acknowledgements

Since this project would not have been possible without the data and support from many people we want to thank everybody who contributed, directly or indirectly, and especially:

- **Steve Lawford**, who very patiently gave answers to most of our questions and without whose support and ideas this project would have never been started, let alone finished.
- **The DOT and BoTS** for providing the data and for very quickly answering our questions.
- **Mehrdad Farzinpour** for maintaining the ENAC aviation database which was very useful during the project.
- **Roseline Bilina** at Cornell University, for very helpful comments on the project results and report.
- And everybody we might have forgot to mention in this list. Thank you!

References

[1] Continental Airlines, awards since 1996. http://www.continental.com/web/en-US/content/company/globalcitizenship/awards.aspx.

[2] Air transport association, 2009 Economic Report, 2009.

[3] American Airlines. http://www.wikinvest.com/stock/American_Airlines_(AMR)#_note-2, 2009.

[4] Barnum, J. W. What prompted airline deregulation 20 years ago? What were the objectives of that deregulation and how were they achieved? http://library.findlaw.com/1988/Sep/1/129304.html, 1988.

[5] Bryant, A. Delta to buy only Boeing jets for 20 years. http://www.nytimes.com/1997/03/21/business/delta-to-buy-only-boeing-jets-for-20-years.html, 1997.

[6] Cobb, J., and Metzger, N. Entry strategy for Southwest airlines. http://www.enac.fr/recherche/leea/SteveLawford/steve_site1_links_student_projects.html, 2009.

[7] Doermer, M., and Lavell, T. Continental Air may reconsider combining with United. http://www.bloomberg.com/apps/news?pid=20601103&sid=aORDoRb8FJqE, 2009.

[8] Fischer, J. W. Aviation finance : FAA reauthorization and related issues. http://assets.opencrs.com/rpts/RL33913_20080421.pdf, 2008.

[9] Guzhva, V. S., and Pagiavlas, N. U.S. commercial airline performance after September 11,2001.

[10] Hamilton, J. D. Time Series Analysis, Princeton : Princeton University Press, 1994.

A Appendix

A.1 ARIMA models (automated outputs)

American		p =	3.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	381.0909	1.056713	-0.415496	0.319092							
se_alpha	123.3027	0.124757	0.175544	0.118374							
t-stat(alpha)	3.090693	8.470207	-2.366904	2.695630							
|AR roots|	0.968016	0.574139	0.574139								
|MA roots|											
R^2	0.917633		SIC	9.328462		max P[Q-stats]	0.629000	P[Q-stats(1)]	0.108000		
D-W	2.015383		JB	1.389676		min P[Q-stats]	0.108000	P[Q-stats(2)]	0.241000		
								P[Q-stats(3)]	0.415000		
Continental		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	431.1574	0.773432									
se_alpha	14.35092	0.070481									
t-stat(alpha)	30.04388	10.97355									
|AR roots|	0.773432										
|MA roots|											
R^2	0.660123		SIC	9.380507		max P[Q-stats]	0.881000	P[Q-stats(1)]	0.751000		
D-W	1.837109		JB	5.986324		min P[Q-stats]	0.029000	P[Q-stats(2)]	0.881000		
								P[Q-stats(3)]	0.029000		
Delta		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	379.8326	0.836018	-0.811370	0.769838	0.426887	0.849721					
se_alpha	26.79139	0.107980	0.131557	0.099230	0.109506	0.097860					
t-stat(alpha)	14.17741	7.742305	-6.167449	7.758092	3.898309	8.683040					
|AR roots|	0.928530	0.928530	0.892909								
|MA roots|	0.921803	0.921803									
R^2	0.905418		SIC	8.799172		max P[Q-stats]	0.935000	P[Q-stats(1)]	0.026000		
D-W	1.898956		JB	9.078859		min P[Q-stats]	0.026000	P[Q-stats(2)]	0.085000		
								P[Q-stats(3)]	0.103000		
Northwest		p =	3.000000	q =	1.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	316.1128	1.381132	-0.807099	0.404651	-0.743569						
se_alpha	90.23796	0.150503	0.191624	0.124109	0.130852						
t-stat(alpha)	3.503102	9.176753	-4.211901	3.260439	-5.682511						
|AR roots|	0.978919	0.642935	0.642935								
|MA roots|	0.743569										
R^2	0.917107		SIC	8.765514		max P[Q-stats]	0.000000	P[Q-stats(1)]	0.000000		
D-W	2.047965		JB	2.857827		min P[Q-stats]	0.000000	P[Q-stats(2)]	0.000000		
								P[Q-stats(3)]	0.000000		
United		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	447.2741	0.677871	-0.740817	0.770504	0.459692	0.826295					
se_alpha	22.55935	0.099997	0.103584	0.093980	0.106240	0.096300					
t-stat(alpha)	19.82655	6.778885	-7.151852	8.198566	4.326906	8.580440					
|AR roots|	0.946938	0.946938	0.859274								
|MA roots|	0.909008	0.909008									
R^2	0.834814		SIC	9.179854		max P[Q-stats]	0.976000	P[Q-stats(1)]	0.096000		
D-W	2.055159		JB	1.985714		min P[Q-stats]	0.096000	P[Q-stats(2)]	0.174000		
								P[Q-stats(3)]	0.261000		
US		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	373.3178	0.882953	-0.949198	0.880449	0.128578	0.951685					
se_alpha	31.05663	0.060565	0.033259	0.060679	0.017726	0.016278					
t-stat(alpha)	12.02055	14.57853	-28.53930	14.50994	7.253694	58.46315					
|AR roots|	0.985331	0.985331	0.906859								
|MA roots|	0.975543	0.975543									
R^2	0.843318		SIC	9.019546		max P[Q-stats]	0.020000	P[Q-stats(1)]	0.001000		
D-W	2.031955		JB	1.525157		min P[Q-stats]	0.001000	P[Q-stats(2)]	0.003000		
								P[Q-stats(3)]	0.003000		
Southwest		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	242.5268	0.812377									
se_alpha	7.462084	0.092025									
t-stat(alpha)	32.50122	8.827793									
|AR roots|	0.812377										
|MA roots|											
R^2	0.660816		SIC	7.252009		max P[Q-stats]	0.999000	P[Q-stats(1)]	0.242000		
D-W	1.796068		JB	0.559358		min P[Q-stats]	0.242000	P[Q-stats(2)]	0.391000		
								P[Q-stats(3)]	0.505000		
Av. fares		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	331.8736	0.837578	-0.820563	0.813071	0.203911	0.933839					
se_alpha	29.42788	0.073421	0.051561	0.063025	0.030858	0.028717					
t-stat(alpha)	11.27752	11.40794	-15.91456	12.90081	6.608099	32.51844					
|AR roots|	0.943393	0.943393	0.913573								
|MA roots|	0.966353	0.966353									
R^2	0.906410		SIC	8.007179		max P[Q-stats]	0.877000	P[Q-stats(1)]	0.054000		
D-W	1.754899		JB	1.075718		min P[Q-stats]	0.054000	P[Q-stats(2)]	0.121000		
								P[Q-stats(3)]	0.210000		

Figure 134: Best ARMA models for filtered data.

American		p =	3.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	378.9510	1.015036	-0.359443	0.307679							
se_alpha	138.4105	0.125448	0.173427	0.117651							
t-stat(alpha)	2.737878	8.091277	-2.072591	2.615185							
|AR roots|	0.971148	0.562868	0.562868								
|MA roots|											
R²	0.918852		SIC	9.294522		max P[Q-stats]	0.742000		P[Q-stats(1)]	0.120000	
D-W	1.997686		JB	1.594823		min P[Q-stats]	0.120000		P[Q-stats(2)]	0.284000	
									P[Q-stats(3)]	0.466000	
Continental		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	434.2499	0.756818									
se_alpha	12.84776	0.073189									
t-stat(alpha)	33.79967	10.34066									
|AR roots|	0.756818										
|MA roots|											
R²	0.632982		SIC	9.301369		max P[Q-stats]	0.684000		P[Q-stats(1)]	0.614000	
D-W	1.847370		JB	3.407983		min P[Q-stats]	0.026000		P[Q-stats(2)]	0.684000	
									P[Q-stats(3)]	0.026000	
Delta		p =	4.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	362.4295	1.419925	-1.121853	1.119001	-0.458336	-0.408297	0.424080	-0.981864			
se_alpha	16.44449	0.124404	0.163324	0.161361	0.112397	0.054459	0.041062	0.044931			
t-stat(alpha)	22.03957	11.41379	-6.868886	6.934751	-4.077817	-7.497300	10.32773	-21.85263			
|AR roots|	0.941583	0.877047	0.877047	0.632820							
|MA roots|	0.997486	0.997486	0.986820								
R²	0.935396		SIC	8.853235		max P[Q-stats]	0.770000		P[Q-stats(1)]	0.037000	
D-W	1.815138		JB	1.689929		min P[Q-stats]	0.037000		P[Q-stats(2)]	0.112000	
									P[Q-stats(3)]	0.219000	
Northwest		p =	2.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	265.0175	0.510736	0.466547	0.386968	-0.584644	-0.794109					
se_alpha	95.22052	0.144505	0.140729	0.111886	0.081452	0.090959					
t-stat(alpha)	2.783197	3.534392	3.315209	3.458577	-7.177800	-8.730397					
|AR roots|	0.984587	0.473851									
|MA roots|	0.997417	0.892281	0.892281								
R²	0.943088		SIC	8.432495		max P[Q-stats]	0.002000		P[Q-stats(1)]	0.002000	
D-W	2.394593		JB	3.759989		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.002000	
									P[Q-stats(3)]	0.000000	
United		p =	3.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	451.2816	1.015383	-0.335958	0.217294							
se_alpha	32.65618	0.112540	0.129389	0.090678							
t-stat(alpha)	13.81918	9.022379	-2.596497	2.396330							
|AR roots|	0.908804	0.488977	0.488977								
|MA roots|											
R²	0.835812		SIC	9.264390		max P[Q-stats]	0.191000		P[Q-stats(1)]	0.002000	
D-W	2.148459		JB	2.908669		min P[Q-stats]	0.002000		P[Q-stats(2)]	0.010000	
									P[Q-stats(3)]	0.027000	
US		p =	3.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	314.7620	0.470643	0.937356	-0.436102	0.424501	-0.980222	-0.437103				
se_alpha	39.46172	0.166977	0.029355	0.145251	0.183268	0.046190	0.146426				
t-stat(alpha)	7.976390	2.818616	31.93220	-3.002413	2.316281	-21.22161	-2.985151				
|AR roots|	0.973354	0.966352	0.463641								
|MA roots|	0.997491	0.970443	0.451549								
R²	0.864778		SIC	9.202308		max P[Q-stats]	0.007000		P[Q-stats(1)]	0.002000	
D-W	1.813794		JB	0.458602		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.001000	
									P[Q-stats(3)]	0.002000	
Southwest		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	247.7865	0.808502									
se_alpha	7.438368	0.092221									
t-stat(alpha)	33.31195	8.767014									
|AR roots|	0.808502										
|MA roots|											
R²	0.657712		SIC	7.295035		max P[Q-stats]	1.000000		P[Q-stats(1)]	0.188000	
D-W	1.754773		JB	0.638795		min P[Q-stats]	0.188000		P[Q-stats(2)]	0.347000	
									P[Q-stats(3)]	0.493000	
Av. fares		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	341.0616	0.829496	-0.815538	0.809642	0.208508	0.930264					
se_alpha	29.44685	0.074183	0.053686	0.063342	0.034310	0.031935					
t-stat(alpha)	11.58228	11.18170	-15.19086	12.78216	6.077257	29.13037					
|AR roots|	0.943006	0.943006	0.910468								
|MA roots|	0.964502	0.964502									
R²	0.906978		SIC	8.076970		max P[Q-stats]	0.536000		P[Q-stats(1)]	0.012000	
D-W	1.887630		JB	1.110087		min P[Q-stats]	0.012000		P[Q-stats(2)]	0.024000	
									P[Q-stats(3)]	0.058000	

Figure 135: Best ARMA models for non filtered data.

American		p =	3.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	378.9510	1.015036	-0.359443	0.307679							
se_alpha	138.4105	0.125448	0.173427	0.117651							
t-stat(alpha)	2.737878	8.091277	-2.072591	2.615185							
|AR roots|	0.971148	0.562868	0.562868								
|MA roots|											
R^2	0.918852		SIC	9.294522		max P[Q-stats]	0.742000		P[Q-stats(1)]	0.120000	
D-W	1.997686		JB	1.594823		min P[Q-stats]	0.120000		P[Q-stats(2)]	0.284000	
									P[Q-stats(3)]	0.466000	
Continental		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	434.2499	0.756818									
se_alpha	12.84776	0.073189									
t-stat(alpha)	33.79967	10.34066									
|AR roots|	0.756818										
|MA roots|											
R^2	0.632982		SIC	9.301369		max P[Q-stats]	0.684000		P[Q-stats(1)]	0.614000	
D-W	1.847370		JB	3.407983		min P[Q-stats]	0.026000		P[Q-stats(2)]	0.684000	
									P[Q-stats(3)]	0.026000	
Delta		p =	4.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	362.4295	1.419925	-1.121853	1.119001	-0.458336	-0.408297	0.424080	-0.981864			
se_alpha	16.44449	0.124404	0.163324	0.161361	0.112397	0.054459	0.041062	0.044931			
t-stat(alpha)	22.03957	11.41379	-6.868886	6.934751	-4.077817	-7.497300	10.32773	-21.85263			
|AR roots|	0.941583	0.877047	0.877047	0.632820							
|MA roots|	0.997486	0.997486	0.986820								
R^2	0.935396		SIC	8.853235		max P[Q-stats]	0.770000		P[Q-stats(1)]	0.037000	
D-W	1.815138		JB	1.689929		min P[Q-stats]	0.037000		P[Q-stats(2)]	0.112000	
									P[Q-stats(3)]	0.219000	
Northwest		p =	2.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	265.0175	0.510736	0.466547	0.386968	-0.584644	-0.794109					
se_alpha	95.22052	0.144505	0.140729	0.111886	0.081452	0.090959					
t-stat(alpha)	2.783197	3.534392	3.315209	3.458577	-7.177800	-8.730397					
|AR roots|	0.984587	0.473851									
|MA roots|	0.997417	0.892281	0.892281								
R^2	0.943088		SIC	8.432495		max P[Q-stats]	0.002000		P[Q-stats(1)]	0.002000	
D-W	2.394593		JB	3.759989		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.002000	
									P[Q-stats(3)]	0.000000	
United		p =	3.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	451.2816	1.015383	-0.335958	0.217294							
se_alpha	32.65618	0.112540	0.129389	0.090678							
t-stat(alpha)	13.81918	9.022379	-2.596497	2.396330							
|AR roots|	0.908804	0.488977	0.488977								
|MA roots|											
R^2	0.835812		SIC	9.264390		max P[Q-stats]	0.191000		P[Q-stats(1)]	0.002000	
D-W	2.148459		JB	2.908669		min P[Q-stats]	0.002000		P[Q-stats(2)]	0.010000	
									P[Q-stats(3)]	0.027000	
US		p =	3.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	314.7620	0.470643	0.937356	-0.436102	0.424501	-0.980222	-0.437103				
se_alpha	39.46172	0.166977	0.029355	0.145251	0.183268	0.046190	0.146426				
t-stat(alpha)	7.976390	2.818616	31.93220	-3.002413	2.316281	-21.22161	-2.985151				
|AR roots|	0.973354	0.966352	0.463641								
|MA roots|	0.997491	0.970443	0.451549								
R^2	0.864778		SIC	9.202308		max P[Q-stats]	0.007000		P[Q-stats(1)]	0.002000	
D-W	1.813794		JB	0.458602		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.001000	
									P[Q-stats(3)]	0.002000	
Southwest		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	247.7865	0.808502									
se_alpha	7.438368	0.092221									
t-stat(alpha)	33.31195	8.767014									
|AR roots|	0.808502										
|MA roots|											
R^2	0.657712		SIC	7.295035		max P[Q-stats]	1.000000		P[Q-stats(1)]	0.188000	
D-W	1.754773		JB	0.638795		min P[Q-stats]	0.188000		P[Q-stats(2)]	0.347000	
									P[Q-stats(3)]	0.493000	
Av. fares		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	341.0616	0.829496	-0.815538	0.809642	0.208508	0.930264					
se_alpha	29.44685	0.074183	0.053686	0.063342	0.034310	0.031935					
t-stat(alpha)	11.58228	11.18170	-15.19086	12.78216	6.077257	29.13037					
|AR roots|	0.943006	0.943006	0.910468								
|MA roots|	0.964502	0.964502									
R^2	0.906978		SIC	8.076970		max P[Q-stats]	0.536000		P[Q-stats(1)]	0.012000	
D-W	1.887630		JB	1.110087		min P[Q-stats]	0.012000		P[Q-stats(2)]	0.024000	
									P[Q-stats(3)]	0.058000	

Figure 136: Best ARMA models for non filtered data.

American		p =	1.000000	q =	1.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	593.2849	-3.535029	0.671194	0.346384							
se_alpha	25.50689	0.648158	0.112386	0.144313							
t-stat(alpha)	23.25979	-5.453966	5.972250	2.400232							
|AR roots|	0.671194										
|MA roots|	0.346384										
R²	0.917817		SIC	9.335251		max P[Q-stats]	0.728000		P[Q-stats(1)]	0.728000	
D-W	1.985877		JB	6.799107		min P[Q-stats]	0.095000		P[Q-stats(2)]	0.155000	
									P[Q-stats(3)]	0.247000	
Continental		p =	2.000000	q =	1.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	548.5347	-2.653741	1.624996	-0.665562	-0.973604						
se_alpha	35.87851	0.749739	0.084832	0.083826	0.020826						
t-stat(alpha)	15.28867	-3.539552	19.15547	-7.939819	-46.74857						
|AR roots|	0.815820	0.815820									
|MA roots|	0.973604										
R²	0.716003		SIC	9.376505		max P[Q-stats]	0.525000		P[Q-stats(1)]	0.006000	
D-W	2.015614		JB	1.977543		min P[Q-stats]	0.006000		P[Q-stats(2)]	0.017000	
									P[Q-stats(3)]	0.040000	
Delta		p =	1.000000	q =	1.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	462.6359	-1.826865	0.705859	0.493513							
se_alpha	26.48296	0.656994	0.085449	0.122755							
t-stat(alpha)	17.46919	-2.780643	8.260588	4.020321							
|AR roots|	0.705859										
|MA roots|	0.493513										
R²	0.907808		SIC	8.832874		max P[Q-stats]	0.378000		P[Q-stats(1)]	0.267000	
D-W	1.983053		JB	0.712007		min P[Q-stats]	0.053000		P[Q-stats(2)]	0.130000	
									P[Q-stats(3)]	0.252000	
Northwest		p =	4.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	553.6550	-2.885838	-0.538583	-0.462829	-0.433866	0.431158	1.402154	1.267999	0.832599		
se_alpha	8.213074	0.210083	0.133407	0.129631	0.123906	0.123118	0.089774	0.126274	0.074805		
t-stat(alpha)	67.41143	-13.73663	-4.037131	-3.570361	-3.501584	3.501986	15.61875	10.04164	11.13032		
|AR roots|	0.978652	0.955920	0.955920	0.482131							
|MA roots|	0.976703	0.923287	0.923287								
R²	0.954315		SIC	8.404671		max P[Q-stats]	0.966000		P[Q-stats(1)]	0.158000	
D-W	1.895132		JB	5.503595		min P[Q-stats]	0.158000		P[Q-stats(2)]	0.334000	
									P[Q-stats(3)]	0.393000	
United		p =	0.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	539.7353	-2.123812	1.052495	0.567186							
se_alpha	15.30335	0.397397	0.107265	0.107329							
t-stat(alpha)	35.26909	-5.344314	9.812121	5.284543							
|AR roots|											
|MA roots|	0.753117	0.753117									
R²	0.836502		SIC	9.192407		max P[Q-stats]	0.131000		P[Q-stats(1)]	0.062000	
D-W	1.858286		JB	1.109041		min P[Q-stats]	0.013000		P[Q-stats(2)]	0.058000	
									P[Q-stats(3)]	0.046000	
US		p =	0.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	465.1703	-2.064267	1.007156	0.611885							
se_alpha	12.42576	0.327728	0.073972	0.053711							
t-stat(alpha)	37.43596	-6.298722	13.61534	11.39222							
|AR roots|											
|MA roots|	0.782231	0.782231									
R²	0.884305		SIC	8.898536		max P[Q-stats]	0.514000		P[Q-stats(1)]	0.514000	
D-W	2.147299		JB	0.198258		min P[Q-stats]	0.011000		P[Q-stats(2)]	0.171000	
									P[Q-stats(3)]	0.275000	
Southwest		p =	0.000000	q =	1.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	198.7432	0.905507	0.542742								
se_alpha	6.502989	0.144790	0.138208								
t-stat(alpha)	30.56182	6.253938	3.927000								
|AR roots|											
|MA roots|	0.542742										
R²	0.733806		SIC	7.120740		max P[Q-stats]	0.990000		P[Q-stats(1)]	0.134000	
D-W	1.722143		JB	1.609048		min P[Q-stats]	0.134000		P[Q-stats(2)]	0.221000	
									P[Q-stats(3)]	0.288000	
Av. fares		p =	0.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	472.9621	-2.552314	1.566027	0.994654							
se_alpha	15.18855	0.309422	0.062126	0.083156							
t-stat(alpha)	31.13939	-8.248660	25.20711	11.96123							
|AR roots|											
|MA roots|	0.997323	0.997323									
R²	0.912782		SIC	7.871818		max P[Q-stats]	0.404000		P[Q-stats(1)]	0.115000	
D-W	2.335391		JB	0.254937		min P[Q-stats]	0.072000		P[Q-stats(2)]	0.072000	
									P[Q-stats(3)]	0.123000	

Figure 137: Best ARMA models for filtered data with the use of a trend variable.

American		p =	3.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	536.0348	-130.2807	1.533481	-1.283428	0.491185	-1.014149	0.637580				
se_alpha	9.954581	12.95449	0.142610	0.195948	0.109847	0.168547	0.154672				
t-stat(alpha)	53.84805	-10.05680	10.75293	-6.549824	4.471542	-6.017027	4.122128				
|AR roots|	0.836501	0.836501	0.701960								
|MA roots|	0.798486	0.798486									
R^2	0.936769		SIC	9.263774		max P[Q-stats]	0.778000		P[Q-stats(1)]	0.040000	
D-W	1.799072		JB	11.11698		min P[Q-stats]	0.040000		P[Q-stats(2)]	0.104000	
									P[Q-stats(3)]	0.141000	
Continental		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	463.9405	-51.42751	0.940090	0.567324							
se_alpha	9.934567	13.91948	0.077616	0.083775							
t-stat(alpha)	46.69962	-3.694642	12.11209	6.771970							
|AR roots|											
|MA roots|	0.753209	0.753209									
R^2	0.739066		SIC	9.350545		max P[Q-stats]	0.000000		P[Q-stats(1)]	0.000000	
D-W	2.121524		JB	2.325072		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.000000	
									P[Q-stats(3)]	0.000000	
Delta		p =	0.000000	q =	3.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	432.7783	-53.97678	1.407549	1.154829	0.477618						
se_alpha	12.39488	15.94027	0.106491	0.119544	0.093838						
t-stat(alpha)	34.91588	-3.386190	13.21755	9.660275	5.089788						
|AR roots|											
|MA roots|	0.812360	0.812360	0.723741								
R^2	0.912791		SIC	9.009704		max P[Q-stats]	0.383000		P[Q-stats(1)]	0.010000	
D-W	2.010130		JB	0.045651		min P[Q-stats]	0.010000		P[Q-stats(2)]	0.032000	
									P[Q-stats(3)]	0.061000	
Northwest		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	501.4677	-88.03957	1.072845	0.635667							
se_alpha	8.857004	12.31818	0.079830	0.079407							
t-stat(alpha)	56.61821	-7.147122	13.43907	8.005211							
|AR roots|											
|MA roots|	0.797287	0.797287									
R^2	0.898308		SIC	8.987061		max P[Q-stats]	0.118000		P[Q-stats(1)]	0.118000	
D-W	1.800074		JB	10.11324		min P[Q-stats]	0.006000		P[Q-stats(2)]	0.006000	
									P[Q-stats(3)]	0.015000	
United		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	506.1110	-77.86129	1.064661	0.673172							
se_alpha	9.798012	13.35465	0.092880	0.092301							
t-stat(alpha)	51.65446	-5.830275	11.46275	7.293217							
|AR roots|											
|MA roots|	0.820470	0.820470									
R^2	0.847939		SIC	9.119888		max P[Q-stats]	0.668000		P[Q-stats(1)]	0.256000	
D-W	1.981516		JB	4.486529		min P[Q-stats]	0.129000		P[Q-stats(2)]	0.344000	
									P[Q-stats(3)]	0.324000	
US		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	428.2453	-64.92823	0.998443	0.605259							
se_alpha	8.702984	12.14641	0.068964	0.052229							
t-stat(alpha)	49.20672	-5.345468	14.47768	11.58860							
|AR roots|											
|MA roots|	0.777984	0.777984									
R^2	0.869696		SIC	9.017450		max P[Q-stats]	0.506000		P[Q-stats(1)]	0.348000	
D-W	2.157277		JB	1.029562		min P[Q-stats]	0.071000		P[Q-stats(2)]	0.295000	
									P[Q-stats(3)]	0.477000	
Southwest		p =	0.000000	q =	1.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	227.4043	14.78263	0.665652								
se_alpha	4.465926	5.285021	0.119385								
t-stat(alpha)	50.91985	2.797081	5.575668								
|AR roots|											
|MA roots|	0.665652										
R^2	0.582633		SIC	7.570481		max P[Q-stats]	0.858000		P[Q-stats(1)]	0.004000	
D-W	1.370375		JB	1.442743		min P[Q-stats]	0.004000		P[Q-stats(2)]	0.017000	
									P[Q-stats(3)]	0.016000	
Av. fares		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	405.9984	-59.84282	1.184851	0.747980							
se_alpha	8.703842	9.674431	0.102280	0.095803							
t-stat(alpha)	46.64588	-6.185668	11.58440	7.807508							
|AR roots|											
|MA roots|	0.864859	0.864859									
R^2	0.898424		SIC	8.024214		max P[Q-stats]	0.122000		P[Q-stats(1)]	0.122000	
D-W	2.113458		JB	4.476622		min P[Q-stats]	0.006000		P[Q-stats(2)]	0.053000	
									P[Q-stats(3)]	0.110000	

Figure 138: Best ARMA models for filtered data with the use of a dummy variable at period 35.

American		p =	3.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	381.0909	1.056713	-0.415496	0.319092							
se_alpha	123.3027	0.124757	0.175544	0.118374							
t-stat(alpha)	3.090693	8.470207	-2.366904	2.695630							
|AR roots|	0.968016	0.574139	0.574139								
|MA roots|											
R^2	0.917633		SIC	9.328462		max P[Q-stats]	0.629000		P[Q-stats(1)]	0.108000	
D-W	2.015383		JB	1.389676		min P[Q-stats]	0.108000		P[Q-stats(2)]	0.241000	
									P[Q-stats(3)]	0.415000	
Continental		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	431.1574	0.773432									
se_alpha	14.35092	0.070481									
t-stat(alpha)	30.04388	10.97355									
|AR roots|	0.773432										
|MA roots|											
R^2	0.660123		SIC	9.380507		max P[Q-stats]	0.881000		P[Q-stats(1)]	0.751000	
D-W	1.837109		JB	5.986324		min P[Q-stats]	0.029000		P[Q-stats(2)]	0.881000	
									P[Q-stats(3)]	0.029000	
Delta		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	379.8326	0.836018	-0.811370	0.769838	0.426887	0.849721					
se_alpha	26.79139	0.107980	0.131557	0.099230	0.109506	0.097860					
t-stat(alpha)	14.17741	7.742305	-6.167449	7.758092	3.898309	8.683040					
|AR roots|	0.928530	0.928530	0.892909								
|MA roots|	0.921803	0.921803									
R^2	0.905418		SIC	8.799172		max P[Q-stats]	0.935000		P[Q-stats(1)]	0.026000	
D-W	1.898956		JB	9.078859		min P[Q-stats]	0.026000		P[Q-stats(2)]	0.085000	
									P[Q-stats(3)]	0.103000	
Northwest		p =	3.000000	q =	1.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	316.1128	1.381132	-0.807099	0.404651	-0.743569						
se_alpha	90.23796	0.150503	0.191624	0.124109	0.130852						
t-stat(alpha)	3.503102	9.176753	-4.211901	3.260439	-5.682511						
|AR roots|	0.978919	0.642935	0.642935								
|MA roots|	0.743569										
R^2	0.917107		SIC	8.765514		max P[Q-stats]	0.000000		P[Q-stats(1)]	0.000000	
D-W	2.047965		JB	2.857827		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.000000	
									P[Q-stats(3)]	0.000000	
United		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	447.2741	0.677871	-0.740817	0.770504	0.459692	0.826295					
se_alpha	22.55935	0.099997	0.103584	0.093980	0.106240	0.096300					
t-stat(alpha)	19.82655	6.778885	-7.151852	8.198566	4.326906	8.580440					
|AR roots|	0.946938	0.946938	0.859274								
|MA roots|	0.909008	0.909008									
R^2	0.834814		SIC	9.179854		max P[Q-stats]	0.976000		P[Q-stats(1)]	0.096000	
D-W	2.055159		JB	1.985714		min P[Q-stats]	0.096000		P[Q-stats(2)]	0.174000	
									P[Q-stats(3)]	0.261000	
US		p =	0.000000	q =	5.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	392.7677	0.843125	0.639845	0.672718	1.078298	0.542170					
se_alpha	10.90251	0.083821	0.030867	0.053818	0.044884	0.079565					
t-stat(alpha)	36.02544	10.05858	20.72897	12.49986	24.02425	6.814178					
|AR roots|											
|MA roots|	0.997421	0.997421	0.927401	0.927401	0.633642						
R^2	0.892250		SIC	8.955838		max P[Q-stats]	0.048000		P[Q-stats(1)]	0.014000	
D-W	1.758185		JB	0.535803		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.048000	
									P[Q-stats(3)]	0.000000	
Southwest		p =	1.000000	q =	0.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	242.5268	0.812377									
se_alpha	7.462084	0.092025									
t-stat(alpha)	32.50122	8.827793									
|AR roots|	0.812377										
|MA roots|											
R^2	0.660816		SIC	7.252009		max P[Q-stats]	0.999000		P[Q-stats(1)]	0.242000	
D-W	1.796068		JB	0.559358		min P[Q-stats]	0.242000		P[Q-stats(2)]	0.391000	
									P[Q-stats(3)]	0.505000	
Av. fares		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	331.8736	0.837578	-0.820563	0.813071	0.203911	0.933839					
se_alpha	29.42788	0.073421	0.051561	0.063025	0.030858	0.028717					
t-stat(alpha)	11.27752	11.40794	-15.91456	12.90081	6.608099	32.51844					
|AR roots|	0.943393	0.943393	0.913573								
|MA roots|	0.966353	0.966353									
R^2	0.906410		SIC	8.007179		max P[Q-stats]	0.877000		P[Q-stats(1)]	0.054000	
D-W	1.754899		JB	1.075718		min P[Q-stats]	0.054000		P[Q-stats(2)]	0.121000	
									P[Q-stats(3)]	0.210000	

Figure 139: Best ARMA models for filtered data (allowing 5 lags for p,q in ARMA(p,q)).

American		p =	2.000000	q =	0.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-3.203286	0.079277	-0.330758								
se_alpha	2.359659	0.122920	0.117912								
t-stat(alpha)	-1.357521	0.644952	-2.805135								
|AR roots|	0.575116	0.575116									
|MA roots|											
R^2	0.120584		SIC	9.280247		max P[Q-stats]	0.656000	P[Q-stats(1)]	0.116000		
D-W	2.024401		JB	0.267391		min P[Q-stats]	0.116000	P[Q-stats(2)]	0.269000		
								P[Q-stats(3)]	0.378000		
Continental		p =	1.000000	q =	1.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-0.444827	0.576384	-0.807316								
se_alpha	1.574538	0.190875	0.147504								
t-stat(alpha)	-0.282513	3.019697	-5.473176								
|AR roots|	0.576384										
|MA roots|	0.807316										
R^2	0.106457		SIC	9.435288		max P[Q-stats]	0.707000	P[Q-stats(1)]	0.280000		
D-W	1.880401		JB	2.018102		min P[Q-stats]	0.056000	P[Q-stats(2)]	0.056000		
								P[Q-stats(3)]	0.098000		
Delta		p =	1.000000	q =	2.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-0.070278	0.722343	-0.689018	-0.634288							
se_alpha	3.021233	0.039780	0.155433	0.154649							
t-stat(alpha)	-0.023261	18.15837	-4.432886	-4.101462							
|AR roots|	0.722343										
|MA roots|	1.212250	0.523232									
R^2	0.455961		SIC	8.593528		max P[Q-stats]	0.117000	P[Q-stats(1)]	0.012000		
D-W	2.188764		JB	1.353955		min P[Q-stats]	0.009000	P[Q-stats(2)]	0.041000		
								P[Q-stats(3)]	0.018000		
Northwest		p =	3.000000	q =	1.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-2.871007	-0.758006	-0.544163	-0.685309	0.733168						
se_alpha	1.091501	0.108750	0.119173	0.092606	0.120323						
t-stat(alpha)	-2.630330	-6.970141	-4.566161	-7.400299	6.093337						
|AR roots|	0.947275	0.850560	0.850560								
|MA roots|	0.733168										
R^2	0.525940		SIC	8.467463		max P[Q-stats]	0.621000	P[Q-stats(1)]	0.208000		
D-W	1.934719		JB	0.507886		min P[Q-stats]	0.075000	P[Q-stats(2)]	0.096000		
								P[Q-stats(3)]	0.192000		
United		p =	1.000000	q =	2.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-1.907227	0.589190	-0.570465	-0.412980							
se_alpha	0.558330	0.125946	0.146020	0.145502							
t-stat(alpha)	-3.415951	4.678132	-3.906761	-2.838309							
|AR roots|	0.589190										
|MA roots|	0.988324	0.417859									
R^2	0.212440		SIC	9.231869		max P[Q-stats]	0.052000	P[Q-stats(1)]	0.003000		
D-W	1.999970		JB	4.964875		min P[Q-stats]	0.003000	P[Q-stats(2)]	0.010000		
								P[Q-stats(3)]	0.025000		
US		p =	1.000000	q =	2.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-2.075390	0.369489	-0.641213	-0.355389							
se_alpha	0.303624	0.145543	0.152563	0.173626							
t-stat(alpha)	-6.835406	2.538687	-4.202940	-2.046857							
|AR roots|	0.369489										
|MA roots|	0.997494	0.356281									
R^2	0.261398		SIC	9.125922		max P[Q-stats]	0.011000	P[Q-stats(1)]	0.002000		
D-W	1.873029		JB	0.582535		min P[Q-stats]	0.000000	P[Q-stats(2)]	0.006000		
								P[Q-stats(3)]	0.007000		
Southwest		p =	0.000000	q =	2.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	0.902712	-0.460501	-0.535261								
se_alpha	0.119819	0.144173	0.141872								
t-stat(alpha)	7.533984	-3.194086	-3.772847								
|AR roots|											
|MA roots|	0.997242	0.536741									
R^2	0.249560		SIC	7.152773		max P[Q-stats]	0.985000	P[Q-stats(1)]	0.080000		
D-W	1.688208		JB	1.560897		min P[Q-stats]	0.080000	P[Q-stats(2)]	0.154000		
								P[Q-stats(3)]	0.192000		
Av. fares		p =	0.000000	q =	3.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-2.576300	0.573635	-0.571335	-0.930671							
se_alpha	0.497620	0.043487	0.032299	0.053338							
t-stat(alpha)	-5.177241	13.19084	-17.68901	-17.44854							
|AR roots|											
|MA roots|	0.979554	0.974729	0.974729								
R^2	0.431063		SIC	7.981590		max P[Q-stats]	0.374000	P[Q-stats(1)]	0.011000		
D-W	2.375552		JB	0.213731		min P[Q-stats]	0.011000	P[Q-stats(2)]	0.027000		
								P[Q-stats(3)]	0.062000		

Figure 140: Best ARMA models for differentiated dependant variable (i.e. best ARIMA(p,1,q) models) in filtered data.

American		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	1.000000
alpha	574.6830	-88.83975	-1.640269	0.882115	0.385701						
se_alpha	13.56196	21.54177	0.589527	0.095828	0.093419						
t-stat(alpha)	42.37462	-4.124070	-2.782348	9.205166	4.128723						
|AR roots|											
|MA roots|	0.621048	0.621048									
R²	0.934305		SIC	9.217373		max P[Q-stats]	0.850000	P[Q-stats(1)]		0.358000	
D-W	2.176079		JB	6.687768		min P[Q-stats]	0.358000	P[Q-stats(2)]		0.595000	
								P[Q-stats(3)]		0.425000	
Continental		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	463.9405	-51.42751	0.940090	0.567324							
se_alpha	9.934567	13.91948	0.077616	0.083775							
t-stat(alpha)	46.69962	-3.694642	12.11209	6.771970							
|AR roots|											
|MA roots|	0.753209	0.753209									
R²	0.739066		SIC	9.350545		max P[Q-stats]	0.000000	P[Q-stats(1)]		0.000000	
D-W	2.121524		JB	2.325072		min P[Q-stats]	0.000000	P[Q-stats(2)]		0.000000	
								P[Q-stats(3)]		0.000000	
Delta		p =	3.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	379.8326	0.836018	-0.811370	0.769838	0.426887	0.849721					
se_alpha	26.79139	0.107980	0.131557	0.099230	0.109506	0.097860					
t-stat(alpha)	14.17741	7.742305	-6.167449	7.758092	3.898309	8.683040					
|AR roots|	0.928530	0.928530	0.892909								
|MA roots|	0.921803	0.921803									
R²	0.905418		SIC	8.799172		max P[Q-stats]	0.935000	P[Q-stats(1)]		0.026000	
D-W	1.898956		JB	9.078859		min P[Q-stats]	0.026000	P[Q-stats(2)]		0.085000	
								P[Q-stats(3)]		0.103000	
Northwest		p =	0.000000	q =	4.000000	d =	1.000000	b =	1.000000	t =	1.000000
alpha	-6.635491	-6.108298	0.196231	-0.602622	-0.945106	-0.271605	0.826250				
se_alpha	0.652090	0.978022	0.033076	0.097917	0.104691	0.094669	6.93E-05				
t-stat(alpha)	-10.17573	-6.245564	5.932793	-6.154409	-9.027595	-2.868982	11916.88				
|AR roots|											
|MA roots|	0.997494	0.997494	0.911266	0.911266							
R²	0.653362		SIC	8.336151		max P[Q-stats]	0.082000	P[Q-stats(1)]		0.050000	
D-W	1.799905		JB	2.102372		min P[Q-stats]	0.002000	P[Q-stats(2)]		0.033000	
								P[Q-stats(3)]		0.004000	
United		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	506.1110	-77.86129	1.064661	0.673172							
se_alpha	9.798012	13.35465	0.092880	0.092301							
t-stat(alpha)	51.65446	-5.830275	11.46275	7.293217							
|AR roots|											
|MA roots|	0.820470	0.820470									
R²	0.847939		SIC	9.119888		max P[Q-stats]	0.668000	P[Q-stats(1)]		0.256000	
D-W	1.981516		JB	4.486529		min P[Q-stats]	0.129000	P[Q-stats(2)]		0.344000	
								P[Q-stats(3)]		0.324000	
US		p =	0.000000	q =	2.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	465.1703	-2.064267	1.007156	0.611885							
se_alpha	12.42576	0.327728	0.073972	0.053711							
t-stat(alpha)	37.43596	-6.298722	13.61534	11.39222							
|AR roots|											
|MA roots|	0.782231	0.782231									
R²	0.884305		SIC	8.898536		max P[Q-stats]	0.514000	P[Q-stats(1)]		0.514000	
D-W	2.147299		JB	0.198258		min P[Q-stats]	0.011000	P[Q-stats(2)]		0.171000	
								P[Q-stats(3)]		0.275000	
Southwest		p =	0.000000	q =	3.000000	d =	0.000000	b =	1.000000	t =	1.000000
alpha	191.2149	-15.54111	1.321151	0.507582	-0.485910	-0.900402					
se_alpha	2.219953	2.749888	0.084972	0.040512	0.032115	0.054744					
t-stat(alpha)	86.13462	-5.651542	15.54803	12.52921	-15.13021	-16.44746					
|AR roots|											
|MA roots|	0.966255	0.966255	0.964392								
R²	0.844654		SIC	6.844582		max P[Q-stats]	0.840000	P[Q-stats(1)]		0.043000	
D-W	1.629990		JB	0.399937		min P[Q-stats]	0.043000	P[Q-stats(2)]		0.075000	
								P[Q-stats(3)]		0.149000	
Av. fares		p =	1.000000	q =	3.000000	d =	0.000000	b =	1.000000	t =	1.000000
alpha	437.6050	-46.32937	-1.011687	0.545707	0.414444	-0.365279	-0.972371				
se_alpha	7.892496	7.343300	0.223626	0.151554	0.037636	0.033268	0.062872				
t-stat(alpha)	55.44571	-6.309066	-4.524019	3.600737	11.01198	-10.97989	-15.46599				
|AR roots|	0.545707										
|MA roots|	0.997463	0.997463	0.977325								
R²	0.934901		SIC	7.829682		max P[Q-stats]	0.855000	P[Q-stats(1)]		0.164000	
D-W	2.026326		JB	0.953751		min P[Q-stats]	0.164000	P[Q-stats(2)]		0.174000	
								P[Q-stats(3)]		0.265000	

Figure 141: Best ARIMA models for filtered data (with possible use of a trend variable and a dummy variable).

American		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	1.000000
alpha	580.3625	-87.85106	-1.650983	0.860514	0.370623						
se_alpha	13.18745	21.04902	0.574834	0.094353	0.092009						
t-stat(alpha)	44.00870	-4.173642	-2.872104	9.120151	4.028100						
|AR roots|											
|MA roots|	0.608788	0.608788									
R²	0.934410		SIC	9.195063		max P[Q-stats]	0.841000		P[Q-stats(1)]	0.268000	
D-W	2.220186		JB	4.988802		min P[Q-stats]	0.268000		P[Q-stats(2)]	0.517000	
									P[Q-stats(3)]	0.358000	
Continental		p =	0.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	466.3134	-51.89094	0.926233	0.548353							
se_alpha	9.208430	12.92411	0.076501	0.083744							
t-stat(alpha)	50.63983	-4.015049	12.10751	6.547973							
|AR roots|											
|MA roots|	0.740508	0.740508									
R²	0.734066		SIC	9.222188		max P[Q-stats]	0.506000		P[Q-stats(1)]	0.212000	
D-W	2.202367		JB	2.128731		min P[Q-stats]	0.046000		P[Q-stats(2)]	0.046000	
									P[Q-stats(3)]	0.098000	
Delta		p =	4.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	0.000000
alpha	362.4295	1.419925	-1.121853	1.119001	-0.458336	-0.408297	0.424080	-0.981864			
se_alpha	16.44449	0.124404	0.163324	0.161361	0.112397	0.054459	0.041062	0.044931			
t-stat(alpha)	22.03957	11.41379	-6.868886	6.934751	-4.077817	-7.497300	10.32773	-21.85263			
|AR roots|	0.941583	0.877047	0.877047	0.632820							
|MA roots|	0.997486	0.997486	0.986820								
R²	0.935396		SIC	8.853235		max P[Q-stats]	0.770000		P[Q-stats(1)]	0.037000	
D-W	1.815138		JB	1.689929		min P[Q-stats]	0.037000		P[Q-stats(2)]	0.112000	
									P[Q-stats(3)]	0.219000	
Northwest		p =	4.000000	q =	3.000000	d =	0.000000	b =	0.000000	t =	1.000000
alpha	556.4219	-2.865983	-0.494535	-0.395445	-0.393415	0.469012	1.427952	1.273337	0.813511		
se_alpha	8.218436	0.210066	0.144094	0.131545	0.124337	0.133880	0.105101	0.148822	0.081167		
t-stat(alpha)	67.70410	-13.64325	-3.432030	-3.006146	-3.164100	3.503215	13.58651	8.556115	10.02267		
|AR roots|	0.986488	0.948656	0.948656	0.528292							
|MA roots|	0.976931	0.912535	0.912535								
R²	0.961141		SIC	8.208349		max P[Q-stats]	0.852000		P[Q-stats(1)]	0.251000	
D-W	1.887232		JB	2.507287		min P[Q-stats]	0.147000		P[Q-stats(2)]	0.494000	
									P[Q-stats(3)]	0.587000	
United		p =	1.000000	q =	2.000000	d =	0.000000	b =	1.000000	t =	0.000000
alpha	522.3056	-88.50288	-0.297339	1.176809	0.780686						
se_alpha	8.217776	11.49746	0.064204	0.090384	0.088512						
t-stat(alpha)	63.55802	-7.697605	-4.631183	13.02007	8.820084						
|AR roots|	0.297339										
|MA roots|	0.883564	0.883564									
R²	0.857745		SIC	9.190899		max P[Q-stats]	0.655000		P[Q-stats(1)]	0.135000	
D-W	1.720066		JB	4.900801		min P[Q-stats]	0.069000		P[Q-stats(2)]	0.159000	
									P[Q-stats(3)]	0.266000	
US		p =	1.000000	q =	2.000000	d =	1.000000	b =	0.000000	t =	0.000000
alpha	-2.784931	0.457360	-0.807764	-0.436972							
se_alpha	1.209650	0.044404	0.160582	0.145174							
t-stat(alpha)	-2.302261	10.29997	-5.030231	-3.009992							
|AR roots|	0.457360										
|MA roots|	1.178538	0.370774									
R²	0.400083		SIC	8.928315		max P[Q-stats]	0.001000		P[Q-stats(1)]	0.000000	
D-W	2.234239		JB	2.103985		min P[Q-stats]	0.000000		P[Q-stats(2)]	0.000000	
									P[Q-stats(3)]	0.000000	
Southwest		p =	0.000000	q =	3.000000	d =	0.000000	b =	1.000000	t =	1.000000
alpha	197.1474	-16.46539	1.327309	0.614065	-0.531715	-0.915340					
se_alpha	2.356047	3.057518	0.093868	0.074627	0.085253	0.043895					
t-stat(alpha)	83.67721	-5.385216	14.14012	8.228480	-6.236935	-20.85293					
|AR roots|											
|MA roots|	0.980219	0.980219	0.952656								
R²	0.843484		SIC	6.886999		max P[Q-stats]	0.717000		P[Q-stats(1)]	0.036000	
D-W	1.706585		JB	1.648904		min P[Q-stats]	0.036000		P[Q-stats(2)]	0.070000	
									P[Q-stats(3)]	0.116000	
Av. fares		p =	1.000000	q =	3.000000	d =	0.000000	b =	1.000000	t =	1.000000
alpha	447.0397	-47.37277	-0.978832	0.545304	0.410159	-0.374553	-0.978057				
se_alpha	8.492168	7.507944	0.226122	0.154522	0.048518	0.032738	0.067188				
t-stat(alpha)	52.64141	-6.309686	-4.328773	3.528977	8.453678	-11.44095	-14.55699				
|AR roots|	0.545304										
|MA roots|	0.997475	0.997475	0.983015								
R²	0.937661		SIC	7.875666		max P[Q-stats]	0.893000		P[Q-stats(1)]	0.146000	
D-W	2.075813		JB	1.341188		min P[Q-stats]	0.146000		P[Q-stats(2)]	0.188000	
									P[Q-stats(3)]	0.308000	

Figure 142: Best ARIMA models for non-filtered data (with possible use of a trend variable and a dummy variable).

A.2 EViews6 Code

This code prepares some pre-treated DOT and some economic data for use by the authors. All EViews6 code for further data treatment, modelling building; and analysis/forecasting was written exclusively by the authors. The first code is a tool to sort the Coupon database in order to get the quarterly average fares and number of passengers per carrier. The second program was used to perform the ARIMA regressions on these data of fares and passengers, and to pick up the best one according to the Schwarz Information Criterion.

```
wfopen complete_coupon_text.wf1


scalar nb_carriers=7
matrix(65,nb_carriers+1) pax
matrix(65,nb_carriers+1) fares


for !k=1 to nb_carriers
for !t=1 to 65
sample paxfares @all if carriercode=!k and time=!t
smpl paxfares
series pax_q=group01(7)
series fares_q=group01(14)
pax(!t,!k)=@sum(pax_q)
series mult=pax_q*fares_q
fares(!t,!k)=@sum(mult)/pax(!t,!k)
delete paxfares
next
next
```

```
for !t=1 to 65
scalar sumpax=0
scalar avfares=0
for !k=1 to nb_carriers
sumpax=sumpax+pax(!t,!k)
avfares=avfares+fares(!t,!k)*pax(!t,!k)
next
pax(!t,nb_carriers+1)=sumpax
fares(!t,nb_carriers+1)=avfares/sumpax
next

fares.write fares.txt
pax.write pax.txt
```

```
wfcreate u 65
'....................................VARIABLES.........................
scalar nb_carriers=8 '{1,2,...8,...}
scalar maxlags=4 '{0,1,...4,...}
scalar maxdifferencing=0 '{0,1,2,...}
scalar trend=0 '{0,1} 1: use of a trend variable
scalar breakdown=0 '{0,1}  1: use of break var. at t=35
scalar filter=1 '{0: no filter; 1:filtered data}
'..................................................................
'..............................MATRIX OF RESULTS.........................
table(10*nb_carriers+1,20) results_bestARMA
matrix(65,nb_carriers) matrix_fares
'..................................................................
'......................CREATION OF continuous SERIES fare!k................
if filter=1 then
matrix_fares.read(t=txt,dropstrings) fares.txt
matrix_fares(1,5)=NA   'missing data from Northwest
else
matrix_fares.read(t=txt,dropstrings) fares2.txt
matrix_fares(12,7)=NA
matrix_fares(13,7)=NA
matrix_fares(12,8)=NA
matrix_fares(13,8)=NA
endif
'....................................................................
```

```
'.....................................COMPUTATIONS...........................
for !k=1 to nb_carriers
vector vect=@columnextract(matrix_fares,!k)
series fare!k
mtos(vect,fare!k)
scalar opt_sic=100
scalar tstat=10
scalar opt_p=0
scalar opt_q=0
scalar opt_d=0
scalar opt_b=0
scalar opt_t=0
scalar opt_qprob_min=0
matrix(50,1) opt_correl
for !p=0 to maxlags
%ar=""
if !p>0 then
for !p2=1 to !p
%ar=%ar+"ar("+@str(!p2)+") "
next
endif
for !q=0 to maxlags
%ma=""
if !q>0 then
for !q2=1 to !q
```

```
%ma=%ma+"ma("+@str(!q2)+") "
next
endif
for !b=0 to breakdown
%break=""
if !b>0 then
series dummy=(@trend>33)
dummy(35)=0.22
%break="dummy "
endif
for !t=0 to trend
%trend=""
if !t>0 then
%trend="@trend "
endif
%arma= %break + %trend +%ar + %ma
for !d=0 to maxdifferencing
%fare="fare!k"
if !d>0 then
%fare="d(fare!k,!d)"
endif
equation fare_!k!p!d!q!b!t.ls {%fare} c {%arma} 'ARMA(p,q)
for !r=1 to !p+!q+!t+!b+1
if fare_!k!p!d!q!b!t.@stderrs(!r,1)<>NA then
if @abs(fare_!k!p!d!q!b!t.@tstats(!r))>1.96 then
```

```
else
tstat=tstat-1
endif
else
tstat=0
endif
next
scalar qprob_min=0
matrix(50,1) temp2
if (filter=1 and (!k!p!d!q!b!t=711001 or !k!p!d!q!b!t=430411
        or !k!p!d!q!t=80001 or  !k!p!d!q!b!t=200210)) then
else
freeze(temp) fare_!k!p!d!q!b!t.correl
for !i=1 to 50
temp2(!i,1)=@val(temp(!i,7))
next
qprob_min=@min (temp2)
endif
if opt_sic>fare_!k!p!d!q!b!t.@schwarz and tstat=10 then   'min{SIC_k}
opt_sic=fare_!k!p!d!q!b!t.@schwarz
opt_p=!p
opt_q=!q
opt_d=!d
opt_b=!b
opt_t=!t
```

```
opt_qprob_min=qprob_min
opt_correl=temp2
endif
tstat=10
delete temp*
next
next
next
next
next
'.................................................................................
'.........................computations of roots...................................
!p=opt_p
!q=opt_q
!d=opt_d
!b=opt_b
!t=opt_t
matrix root_ar
matrix root_ma
if maxlags>0 and (opt_p>0 or opt_q>0) then
fare_!k!p!d!q!b!t.arma(type=root,save=root)
close fare_!k!p!d!q!b!t
endif
fare_!k!p!d!q!b!t.makeresids residuals
freeze(residu) residuals.stats
```

```
'.........................................................................
'..................................RESULTS BEST_ARMA..........................
results_bestARMA((!k-1)*10+1,4)=!p 'p_opt
results_bestARMA((!k-1)*10+1,6)=!q 'q_opt
results_bestARMA((!k-1)*10+1,8)=!d 'q_opt
results_bestARMA((!k-1)*10+1,10)=!b 'q_opt
results_bestARMA((!k-1)*10+1,12)=!t 'q_opt
for !r=1 to !p+!q+!b+!t+1
results_bestARMA((!k-1)*10+2,!r+1)=fare_!k!p!d!q!b!t.@coefs(!r,1) 'apha
results_bestARMA((!k-1)*10+3,!r+1)=fare_!k!p!d!q!b!t.@stderrs(!r,1) 'se_apha
if fare_!k!p!d!q!b!t.@stderrs(!r,1)>0 then
results_bestARMA((!k-1)*10+4,!r+1)=fare_!k!p!d!q!b!t.@tstats(!r) 't-stat(alpha)
endif
next
for !r=1 to !p
results_bestARMA((!k-1)*10+5,!r+1)=root_ar(!r,3) 'AR roots
next
for !r=1 to !q
results_bestARMA((!k-1)*10+6,!r+1)=root_ma(!r,3) 'MA roots
next
results_bestARMA((!k-1)*10+7,2)=fare_!k!p!d!q!b!t.@r2 'R
results_bestARMA((!k-1)*10+7,5)=opt_sic 'SIC
results_bestARMA((!k-1)*10+8,2)=fare_!k!p!d!q!b!t.@dw 'Durbin-Watson
results_bestARMA((!k-1)*10+8,5)=residu(14,2) 'Jarque-Bera
!i=1
```

```
while opt_correl(!i,1)=NA
!i=!i+1
wend
results_bestARMA((!k-1)*10+7,11)=opt_correl(!i,1) 'Q_stat(1)
results_bestARMA((!k-1)*10+8,11)=opt_correl(!i+1,1) 'Q_stat(2)
results_bestARMA((!k-1)*10+9,11)=opt_correl(!i+2,1) 'Q_stat(3)
results_bestARMA((!k-1)*10+7,8)=@max(opt_correl) 'Q_stat max
results_bestARMA((!k-1)*10+8,8)=@min(opt_correl) 'Q_stat minimum
'.........................................................................................
'.....................................LEGENDS and FORMATTING..............................
results_bestARMA(1,1)="American"
results_bestARMA(11,1)="Continental"
results_bestARMA(21,1)="Delta"
results_bestARMA(31,1)="Northwest"
results_bestARMA(41,1)="United"
results_bestARMA(51,1)="US"
results_bestARMA(61,1)="Southwest"
results_bestARMA(71,1)="Av. fares"
!coord=(!k-1)*10+2
!coord2=(!k-1)*10+1
!coord3=(!k-1)*10+5
!coord4=(!k-1)*10+7
!coord5=(!k-1)*10+10
results_bestARMA.setlines(!coord2) +t
results_bestARMA.setlines(!coord) +t
```

```
results_bestARMA.setlines(!coord3) +t
results_bestARMA.setlines(!coord4) +t
results_bestARMA.setlines(a) +r
results_bestARMA.setfillcolor(!coord2) @rgb(50,255,0)
results_bestARMA.settextcolor(!coord,2) @rgb(0,0,255)
results_bestARMA.setwidth(2:20) 9
results_bestARMA.setheight(!coord5) 0
results_bestARMA(!coord,1)="alpha"
results_bestARMA(!coord+1,1)="se_alpha"
results_bestARMA(!coord+2,1)="t-stat(alpha)"
results_bestARMA(!coord+3,1)="|AR roots|"
results_bestARMA(!coord+4,1)="|MA roots|"
results_bestARMA(!coord+5,1)="R"
results_bestARMA(!coord+5,4)="SIC"
results_bestARMA(!coord+6,1)="D-W"
results_bestARMA(!coord+6,4)="JB"
results_bestARMA(!coord+5,6)="         max P[Q-stats]"
results_bestARMA(!coord+6,6)="          min P[Q-stats]"
results_bestARMA(!coord+5,9)="            P[Q-stats(1)]"
results_bestARMA(!coord+6,9)="            P[Q-stats(2)]"
results_bestARMA(!coord+7,9)="            P[Q-stats(3)]"
results_bestARMA(!coord-1,3)="      p ="
results_bestARMA(!coord-1,5)="      q ="
results_bestARMA(!coord-1,7)="      d ="
results_bestARMA(!coord-1,9)="      b ="
```

```
results_bestARMA(!coord-1,11)="    t ="
'.............................................................................
'............................FORECAST.........................................
pagestruct(end=73)
series dummy=(@trend>33)
dummy(35)=0.22
rename fare_!k!p!d!q!b!t carrier_!k
smpl 66 73
carrier_!k.forecast ffare!k ffare_se!k
smpl 58 65
carrier_!k.forecast ffare2!k ffare2_se!k
smpl @all
graph gr!k.line ffare2!k ffare!k fare!k ffare!k+ffare_se!k ffare!k-ffare_se!k
'gr!k.legend -display
gr!k.name(3)
gr!k.name(4)
gr!k.name(5)
gr!k.name(6)
gr!k.name(7)
gr!k.name(1) forecast!k1
gr!k.name(2) forecast!k2
%gr=%gr+" gr" + @str(!k)
%f1=%f1+" ffare" + @str(!k)
%f2=%f2+" ffare2" + @str(!k)
group fare_g1 {%f1}
```

```
group fare_g2 {%f2}
'........................................................................................
'.........................................................................................
delete opt*
delete residu*
delete qprob_min
delete tstat
delete root_*
delete vect
'........................................................................................
'........................................................................................
next
'.........................................................................................
'.........................................................................................
graph forecast.merge {%gr}
matrix matrix_faresf=@convert(fare_g1)
matrix matrix_faresf2=@convert(fare_g2)
delete gr*
delete ffare*
delete fare_*
show forecast
show results_bestarma
```

A.3 Excel VBA-code for Form 41 data treatment

The following VBA programs were written exclusively by the authors, and were used to treat and convert the Form 41 data into a format that could be used for modeling in EViews6.

```
Sub spaltensort()


Dim i As Integer, w As Integer
Dim letztespalte As Integer, namespalte As Integer, _
yearspalte As Integer, quarterspalte As Integer
Dim name As String



For w = 1 To Worksheets.Count

    letztespalte = 0

    'Carriername verschieben
    For i = 1 To 99
        If Worksheets(w).Cells(1, i).Value = "UNIQUE_CARRIER_NAME" Then
            namespalte = i
        End If
    Next
    If Not namespalte = 1 Then
        Worksheets(w).Columns(namespalte).Cut
        Worksheets(w).Columns(1).Insert Shift:=xlToRight
```

```
End If
namespalte = 3

'letzte Spalte finden
For i = 1 To 99
    If Worksheets(w).Cells(1, i).Value = "" Then
        letztespalte = i - 1
    End If
Next i

' Quarter verschieben
For i = 1 To 99
    If Worksheets(w).Cells(1, i).Value = "QUARTER" Then
        quarterspalte = i
    End If
Next
If Not quarterspalte = 1 Then
    Worksheets(w).Columns(quarterspalte).Cut
    Worksheets(w).Columns(1).Insert Shift:=xlToRight
End If
quarterspalte = 2

'Jahr verschieben
For i = 1 To 99
    If Worksheets(w).Cells(1, i).Value = "YEAR" Then
```

```
            yearspalte = i
        End If
    Next
    If Not yearspalte = 1 Then
        Worksheets(w).Columns(yearspalte).Cut
        Worksheets(w).Columns(1).Insert Shift:=xlToRight
    End If
    yearspalte = 1


'worksheet next:
Next

--------------------------------------

End Sub

Sub employquarter()

Dim w As Integer, i As Integer, z As Integer

For w = 1 To Worksheets.Count
    If Left(Worksheets(w).name, 4) = "EMPL" Then
         For i = 1 To 99
            If Worksheets(w).Cells(1, i).Value = "" Then
```

```
                Worksheets(w).Cells(1, i).Value = "QUARTER"
                For z = 2 To Worksheets(w).Cells(Rows.Count, 1).End(xlUp).row
                    Worksheets(w).Cells(z, i).Value = "1"
                Next
            End If
        Next
    End If
Next

End Sub

--------------------------------------

Sub entfquarter()

For w = 1 To Worksheets.Count
    If Left(Worksheets(w).name, 4) = "EMPL" Then
            Worksheets(w).Columns("J:CU").ClearContents
    End If
Next


End Sub

--------------------------------------
```

```
Sub headercheck()

Dim w As Integer
Dim i As Integer, breite As Integer


For w = 2 To Worksheets.Count
    If Left(Worksheets(w).name, 4) = Left(Worksheets(w - 1).name, 4) Then
        breite = 1
        Do While Worksheets(w).Cells(1, breite).Value <> ""
            breite = breite + 1
        Loop
        For i = 1 To breite
            If Worksheets(w).Cells(1, i).Value <> _
Worksheets(w - 1).Cells(1, i).Value Then
                Debug.Print Worksheets(w).name; _
Worksheets(w).Cells(1, i).Value; Worksheets(w - 1).Cells(1, i).Value
            End If
        Next
    End If
Next

End Sub
```

```
Sub identitycheck()

Dim lastrow As Integer, i As Integer


lastrow = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row

For i = 2 To lastrow
    If Worksheets(1).Cells(i, 3).Value <> Worksheets(1).Cells(i, 29).Value _
    Or Worksheets(1).Cells(i, 3).Value <> Worksheets(1).Cells(i, 38).Value _
    Or Worksheets(1).Cells(i, 2).Value <> Worksheets(1).Cells(i, 28).Value _
    Or Worksheets(1).Cells(i, 2).Value <> Worksheets(1).Cells(i, 37).Value _
    Or Worksheets(1).Cells(i, 1).Value <> Worksheets(1).Cells(i, 27).Value _
    Or Worksheets(1).Cells(i, 1).Value <> Worksheets(1).Cells(i, 36).Value Then
        Debug.Print i
    End If
Next
End Sub
```

```
Sub zusammenfassen()
```

```
Dim i As Integer, w As Integer, z As Integer, lastrow As Integer, y As Integer
Dim letztespalte As Integer, namespalte As Integer, yearspalte As Integer,_
 quarterspalte As Integer
Dim name As String, year As Integer, quarter As Integer
Dim gesuchtespalte As String

For w = 2 To Worksheets.Count
    Debug.Print Worksheets(w).name
    'letzte freie zeile finden
    lastrow = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row

    For i = 2 To lastrow
        If Worksheets(w).Cells(i, 1).Value <> "" _
And Worksheets(w).Cells(i, 2).Value <> "" Then
            z = 2
            name = Worksheets(w).Cells(i, 3).Value
            quarter = Worksheets(w).Cells(i, 2).Value
            year = Worksheets(w).Cells(i, 1).Value

            If name = "" Then
                z = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row + 1
                'namespalte = Worksheets(w).Cells(1, 4).Value

                If Left(Worksheets(w).name, 4) = "EMPL" Then
                    Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
```

```
                Worksheets(w).Cells(i, 3)).Copy Destination:=Worksheets(1).Cells(z, 1)
                Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
                Worksheets(w).Cells(i, 9)).Cut Destination:=Worksheets(1).Cells(z, 27)
            ElseIf Left(Worksheets(w).name, 4) = "ACCR" Then
                Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
                Worksheets(w).Cells(i, 3)).Copy Destination:=Worksheets(1).Cells(z, 1)
                Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
                Worksheets(w).Cells(i, 9)).Cut Destination:=Worksheets(1).Cells(z, 36)
            Else
                Debug.Print Worksheets(w).name
                MsgBox ("Fehler!")
            End If
        End If

        Do Until z > lastrow Or Worksheets(1).Cells(z, 2).Value = quarter_
And Worksheets(1).Cells(z, 1).Value = year And Worksheets(1).Cells(z, 3).Value = name
            Do While Worksheets(1).Cells(z, 1).Value <> year And z < lastrow + 1
                z = z + 1
            Loop

            Do While Worksheets(1).Cells(z, 2).Value <> quarter And z < lastrow + 1
                z = z + 1
            Loop

            Do While Worksheets(1).Cells(z, 3).Value <> name And z < lastrow + 1
```

```
                z = z + 1
            Loop

            If Left(Worksheets(w).name, 4) = "EMPL" Then
                Do While Worksheets(1).Cells(z, 27).Value <> "" And z < lastrow + 1
                    z = z + 1
                Loop
            ElseIf Left(Worksheets(w).name, 4) = "ACCR" Then
                Do While Worksheets(1).Cells(z, 36).Value <> "" And z < lastrow + 1
                    z = z + 1
                Loop
            End If
        Loop


        'Debug.Print z, Worksheets(1).Cells(z, 3).Value, name, year, quarter

        If Worksheets(1).Cells(z, 2).Value <> quarter Or _
Worksheets(1).Cells(z, 1).Value <> year Or Worksheets(1).Cells(z, 3).Value <> name Then
            If z <> lastrow + 1 Then
                Debug.Print z, Worksheets(1).Cells(z, 2).Value, _
quarter, Worksheets(1).Cells(z, 1).Value, year, Worksheets(1).Cells(z, 3).Value, name
            Else
                z = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row + 1
                'namespalte = Worksheets(w).Cells(1, 4).Value
```

```
                    If Left(Worksheets(w).name, 4) = "EMPL" Then
                        Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
Worksheets(w).Cells(i, 3)).Copy Destination:=Worksheets(1).Cells(z, 1)
                        Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
Worksheets(w).Cells(i, 9)).Cut Destination:=Worksheets(1).Cells(z, 27)
                    ElseIf Left(Worksheets(w).name, 4) = "ACCR" Then
                        Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
Worksheets(w).Cells(i, 3)).Copy Destination:=Worksheets(1).Cells(z, 1)
                        Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
Worksheets(w).Cells(i, 9)).Cut Destination:=Worksheets(1).Cells(z, 36)
                    Else
                        Debug.Print Worksheets(w).name
                        MsgBox ("Fehler!")
                    End If
                End If
                'Debug.Print z, Worksheets(1).Cells(z, 2).Value, _
quarter, Worksheets(1).Cells(z, 1).Value, year, Worksheets(1).Cells(z, 3).Value, name
                'MsgBox ("Error. nicht richtige zeile erwischt. " & _
                name & " nicht gleich " & Worksheets(1).Cells(z, 3).Value)
            ElseIf name <> "" Then
                If Left(Worksheets(w).name, 4) = "EMPL" Then
                    Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
                    Worksheets(w).Cells(i, 9)).Cut Destination:=Worksheets(1).Cells(z, 27)
                ElseIf Left(Worksheets(w).name, 4) = "ACCR" Then
```

```
                Worksheets(w).Range(Worksheets(w).Cells(i, 1), _
                Worksheets(w).Cells(i, 9)).Cut Destination:=Worksheets(1).Cells(z, 36)
            Else
                Debug.Print Worksheets(w).name
                MsgBox ("Error")
            End If
        End If
    End If
    If w = Worksheets.Count And i = lastrow Then
        Debug.Print i, lastrow, Worksheets.Count, w
        Exit Sub
    End If
Next
Next

End Sub

--------------------------------------

Sub financialzusammenfuegen()

Dim breite As Integer, hoehe As Integer
```

```
For w = 2 To Worksheets.Count
    If Left(Worksheets(w).name, 4) = "FINA" Then
        breite = 1
        hoehe = 1

        Do While Worksheets(w).Cells(1, breite).Value <> ""
            breite = breite + 1
        Loop
        hoehe = Worksheets(w).Cells(Rows.Count, 1).End(xlUp).row

        Worksheets(w).Range(Worksheets(w).Cells(2, 1), _
Worksheets(w).Cells(hoehe, breite - 1)).Cut _
Destination:=Worksheets(1).Cells(Worksheets(1)._
Cells(Rows.Count, 1).End(xlUp).row + 1, 1)
    End If
Next

End Sub

--------------------------------------

Sub spaltensort()

Dim w As Integer
```

```
For w = 1 To Worksheets.Count
    Worksheets(w).Range("A1:ZZ999").Sort Key1:=Worksheets(w).Range("A1"),_
     Order1:=xlAscending, Orientation:=xlColumns
Next

End Sub

--------------------------------------

Sub spaltsort2()

    With ActiveWorkbook.Worksheets(3).Sort
        .SetRange Range("A1:Z350")
        .Header = xlYes
        .MatchCase = False
        .Orientation = xlLeftToRight
        .SortMethod = xlPinYin
        .Apply
    End With
End Sub

--------------------------------------

Sub sum_regions()
```

```
Dim i As Integer, s As Integer
Dim row As Integer

row = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row
i = 2

Do Until i > row
    If Cells(i, 1).Value = Cells(i + 1, 1).Value And Cells(i, 2).Value = _
    Cells(i + 1, 2).Value And Cells(i, 3).Value = Cells(i + 1, 3).Value Then
        For s = 4 To 26
            Cells(i, s).Value = Cells(i, s).Value + Cells(i + 1, s).Value
        Next
        For s = 30 To 35
            Cells(i, s).Value = Cells(i, s).Value + Cells(i + 1, s).Value
        Next
        Cells(i, 39).Value = Cells(i, 39).Value + Cells(i + 1, 39).Value

        Rows(i + 1).Delete
        row = row - 1
        i = i - 1
    End If
    i = i + 1
    Debug.Print i
Loop
```

```
End Sub

--------------------------------------

Sub identitycheck2()


Dim lastrow As Integer, i As Integer


lastrow = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row

For i = 2 To lastrow
    If Worksheets(1).Cells(i, 3).Value <> Worksheets(1).Cells(i, 29).Value _
    And Worksheets(1).Cells(i, 29).Value <> "" _
    Or Worksheets(1).Cells(i, 3).Value <> Worksheets(1).Cells(i, 38).Value _
    And Worksheets(1).Cells(i, 38).Value <> "" _
    Or Worksheets(1).Cells(i, 2).Value <> Worksheets(1).Cells(i, 28).Value _
    And Worksheets(1).Cells(i, 28).Value <> "" _
    Or Worksheets(1).Cells(i, 2).Value <> Worksheets(1).Cells(i, 37).Value _
    And Worksheets(1).Cells(i, 37).Value <> "" _
    Or Worksheets(1).Cells(i, 1).Value <> Worksheets(1).Cells(i, 27).Value _
    And Worksheets(1).Cells(i, 27).Value <> "" _
    Or Worksheets(1).Cells(i, 1).Value <> Worksheets(1).Cells(i, 36).Value _
    And Worksheets(1).Cells(i, 36).Value <> "" Then
```

```
        Debug.Print i,
    End If
Next

End Sub

--------------------------------------

Sub cellformat()

Dim lastrow As Integer
Dim s As Integer, i As Integer


lastrow = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row

For s = 4 To 26
    For i = 2 To lastrow
        Worksheets(1).Cells(i, s).Value = CDbl(Worksheets(1).Cells(i, s).Value) / 100
    Next
Next

For i = 2 To lastrow
    Worksheets(1).Cells(i, 39).Value = CDbl(Worksheets(1).Cells(i, 39).Value) / 100
Next
```

```
End Sub

--------------------------------------

Sub create_ws()

For i = 4 To 26
    Worksheets.Add after:=Worksheets(Worksheets.Count)
    Worksheets(Worksheets.Count).name = Worksheets(1).Cells(1, i).Value
Next

For i = 30 To 36
    Worksheets.Add after:=Worksheets(Worksheets.Count)
    Worksheets(Worksheets.Count).name = Worksheets(1).Cells(1, i).Value
Next

Worksheets.Add after:=Worksheets(Worksheets.Count)
Worksheets(Worksheets.Count).name = Worksheets(1).Cells(1, 39).Value



End Sub
```

```
--------------------------------------

Sub singlesheets_financial()

Dim i As Integer, c As Integer, s As Integer, quarter As Integer
Dim name As String

For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row - 1
    quarter = (Worksheets(1).Cells(i, 1).Value - 1990) * 4 + Worksheets(1).Cells(i, 2).Value
    If Worksheets(1).Cells(i, 3).Value = "American Airlines Inc." Then
        c = 2
    ElseIf Worksheets(1).Cells(i, 3).Value = "Continental Air Lines Inc." Then
        c = 3
    ElseIf Worksheets(1).Cells(i, 3).Value = "Delta Air Lines Inc." Then
        c = 4
    ElseIf Worksheets(1).Cells(i, 3).Value = "Northwest Airlines Inc." Then
        c = 5
    ElseIf Worksheets(1).Cells(i, 3).Value = "Southwest Airlines Co." Then
        c = 6
    ElseIf Worksheets(1).Cells(i, 3).Value = "United Air Lines Inc." Then
        c = 7
    ElseIf Worksheets(1).Cells(i, 3).Value = "US Airways Inc." Then
        c = 8
    Else
        c = 0
```

```
    End If

    If c <> 0 Then

        For s = 4 To 26
            name = Worksheets(1).Cells(1, s).Value
            Worksheets(name).Cells(quarter + 1, c).Value = Worksheets(1).Cells(i, s).Value
        Next

        For s = 30 To 35
            name = Worksheets(1).Cells(1, s).Value
            Worksheets(name).Cells(quarter + 1, c).Value = Worksheets(1).Cells(i, s).Value
        Next

        Worksheets("ACCR_SALARIES").Cells(quarter + 1, c).Value = _
        Worksheets(1).Cells(i, 39).Value

    End If

Next

End Sub

--------------------------------------
```

```
Sub inflate_financial()

Dim w As Integer, i As Integer, s As Integer
Dim quarter As Integer, row As Integer

w = 3

Do Until Worksheets(w).name = "GENERAL_MANAGE"
    For i = 2 To 68
        quarter = Worksheets(w).Cells(i, 1)
        row = 2 + (3 * (quarter - 1))
        For s = 2 To 8
            Worksheets(w).Cells(i, s) = Worksheets(w).Cells(i, s).Value * _
            Worksheets("Inflation").Cells(row, 8)
        Next
    Next
    w = w + 1
    Worksheets(w).Activate
Loop


End Sub

--------------------------------------
```

```
Sub delete1990()

For w = 3 To Worksheets.Count
    Worksheets(w).Range("B2:H13").Delete Shift:=xlUp
    Worksheets(w).Range("A69:A80").ClearContents
Next

End Sub

Sub quarterize_employees()

Dim w As Worksheet
Dim i As Integer, s As Integer, z As Integer

Dim a As Integer



Set w = Worksheets("TOTAL")
i = 3

w.Columns(2).Insert
w.Cells(1, 2).Value = "QUARTER"

Do Until w.Cells(i, 1).Value = ""
```

```
    If w.Cells(i, 1).Value <> w.Cells(i + 1, 1).Value Then
        w.Rows(i).Insert
        w.Rows(i).Insert
        w.Rows(i).Insert
        For s = 3 To 9
            a = (w.Cells(i + 3, s).Value - w.Cells(i - 1, s).Value) / 4
            For z = 0 To 2
                w.Cells(i + z, s).Value = w.Cells(i - 1, s).Value + (z + 1) * a
            Next
        Next
        For z = 0 To 2
            w.Cells(i + z, 1).Value = w.Cells(i - 1, 1).Value
            w.Cells(i + z, 2).Value = z + 2
        Next
        w.Cells(i - 1, 2).Value = 1
    End If
    i = i + 4
Loop


End Sub

--------------------------------------
```

```
Sub quarterize_pilots()

Dim w As Worksheet
Dim i As Integer, s As Integer, z As Integer

Dim a As Integer



Set w = Worksheets("PILOTS_COPILOTS")
i = 3

w.Columns(2).Insert
w.Cells(1, 2).Value = "QUARTER"

Do Until w.Cells(i, 1).Value = ""

    If w.Cells(i, 1).Value <> w.Cells(i + 1, 1).Value Then
        w.Rows(i).Insert
        w.Rows(i).Insert
        w.Rows(i).Insert
        For s = 3 To 9
            a = (w.Cells(i + 3, s).Value - w.Cells(i - 1, s).Value) / 4
            For z = 0 To 2
                w.Cells(i + z, s).Value = w.Cells(i - 1, s).Value + (z + 1) * a
```

```
            Next
        Next
        For z = 0 To 2
            w.Cells(i + z, 1).Value = w.Cells(i - 1, 1).Value
            w.Cells(i + z, 2).Value = z + 2
        Next
        w.Cells(i - 1, 2).Value = 1
    End If
    i = i + 4
Loop

End Sub

--------------------------------------

Sub fueldatain1()

Dim breite As Integer, hoehe As Integer

For w = 2 To Worksheets.Count
        breite = 1
        hoehe = 1

        Do While Worksheets(w).Cells(1, breite).Value <> ""
            breite = breite + 1
```

```
        Loop
        hoehe = Worksheets(w).Cells(Rows.Count, 1).End(xlUp).row

        Worksheets(w).Range(Worksheets(w).Cells(2, 1), _
Worksheets(w).Cells(hoehe, breite - 1)).Cut _
Destination:=Worksheets(1).Cells(Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row + 1, 1)
Next

End Sub

--------------------------------------

Sub monthtoquarter()

Dim i As Integer, s As Integer
Dim row As Integer

row = Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row
i = 2

Do Until i > row
    If Cells(i, 1).Value = Cells(i + 1, 1).Value And _
    Cells(i, 2).Value = Cells(i + 1, 2).Value _
    And Cells(i, 4).Value = Cells(i + 1, 4).Value Then
        For s = 6 To 13
```

```
            Cells(i, s).Value = Cells(i, s).Value + _
            Cells(i + 1, s).Value
        Next
        Rows(i + 1).Delete
        row = row - 1
        i = i - 1
    End If
    i = i + 1
Loop

End Sub

--------------------------------------

Sub zerotest()

Dim i As Integer, s As Integer

For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row
    Debug.Print i
    For s = 6 To 9
        If Right(Cells(i, s).Value, 2) <> "00" Then
            mbox = ("Zeile" & i & ", Spalte" & s)
        End If
    Next
```

```
Next



End Sub

--------------------------------------

Sub convertdbl()

Dim i As Integer, s As Integer

For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row
    Debug.Print i
    For s = 6 To 9
        Cells(i, s).Value = CDbl(Cells(i, s).Value) / 100
    Next
Next



End Sub

--------------------------------------
```

```
Sub carriercodetest()

Dim i As Integer, s As Integer

For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row - 1
    Debug.Print i
    For s = 6 To 9
        If Cells(i, 4).Value = Cells(i + 1, 4).Value Then
            mbox = ("Zeile" & i & ", Spalte" & s)
        End If
    Next
Next

End Sub

--------------------------------------

Sub inflate()

Dim year As Integer, month As Integer
Dim row As Integer
Dim i As Integer


For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row
```

```
    year = Worksheets(1).Cells(i, 1).Value
    month = Worksheets(1).Cells(i, 3).Value
    row = (year - 1993) * 12 + month + 1
    Worksheets(1).Cells(i, 10).Value = Worksheets(1).Cells(i, 8).Value * _
    Worksheets("Inflation").Cells(row, 4)
    Worksheets(1).Cells(i, 11).Value = Worksheets(1).Cells(i, 9).Value * _
    Worksheets("Inflation").Cells(row, 4)
    Worksheets(1).Cells(i, 12).Value = Worksheets(1).Cells(i, 8).Value * _
    Worksheets("Inflation").Cells(row, 5)
    Worksheets(1).Cells(i, 13).Value = Worksheets(1).Cells(i, 9).Value * _
    Worksheets("Inflation").Cells(row, 5)
Next

End Sub

--------------------------------------

Sub inflate2()

Dim year As Integer, quarter As Integer
Dim row As Integer
Dim i As Integer


For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row
```

```
    year = Worksheets(1).Cells(i, 1).Value
    quarter = Worksheets(1).Cells(i, 2).Value
    row = (year - 1993) * 12 + (quarter - 1) * 3 + 2
    Worksheets(1).Cells(i, 13).Value = Worksheets(1).Cells(i, 7).Value * _
    Worksheets("Inflation").Cells(row, 8).Value
    Worksheets(1).Cells(i, 14).Value = Worksheets(1).Cells(i, 8).Value * _
    Worksheets("Inflation").Cells(row, 8).Value
Next

End Sub

--------------------------------------

Sub singlesheets()

Dim i As Integer, c As Integer, x As Integer, quarter As Integer

For i = 2 To Worksheets(1).Cells(Rows.Count, 1).End(xlUp).row - 1
    quarter = Worksheets(1).Cells(i, 3).Value
    If Worksheets(1).Cells(i, 4).Value = "AA" Then
        c = 2
    ElseIf Worksheets(1).Cells(i, 4).Value = "CO" Then
        c = 3
    ElseIf Worksheets(1).Cells(i, 4).Value = "DL" Then
        c = 4
```

```
    ElseIf Worksheets(1).Cells(i, 4).Value = "NW" Then
        c = 5
    ElseIf Worksheets(1).Cells(i, 4).Value = "WN" Then
        c = 6
    ElseIf Worksheets(1).Cells(i, 4).Value = "UA" Then
        c = 7
    ElseIf Worksheets(1).Cells(i, 4).Value = "US" Then
        c = 8
    Else
        c = 0
    End If

    If c <> 0 Then
        Worksheets("TDOMT_GALLONS").Cells(quarter - 27, c).Value = _
        Worksheets(1).Cells(i, 6).Value
        Worksheets("TOTAL_GALLONS").Cells(quarter - 27, c).Value = _
        Worksheets(1).Cells(i, 7).Value
        Worksheets("TDOMT_COST").Cells(quarter - 27, c).Value = _
        Worksheets(1).Cells(i, 10).Value
        Worksheets("TOTAL_COST").Cells(quarter - 27, c).Value = _
        Worksheets(1).Cells(i, 11).Value
    End If

Next
```

```
End Sub
```

A.4 Python Code

The two extended Python programs for preliminary data treatment in this project were developed by Steve Lawford.

```
"""

Parse complete_coupon dictionary to text.
(for project with Anthony/Maximilian)

Steve Lawford: 20 November 2009

"""

import cPickle,time
from safe_cPickle import *

# Set oneway_flag=1 for one-way flights only (=0 for return flights)
oneway_flag=0

if oneway_flag:
    data_path='C:/Project/data/complete_coupon_oneway/'
    filename='complete_coupon_oneway'
else:
    data_path='C:/Project/data/complete_coupon/'
    filename='complete_coupon'
```

```
# cPickle load individual coupon_full_x_x dictionaries

t0=time.time(); print 'Loading dictionary: '+filename
c_x_x=open(data_path+filename,'r')
dt=cPickle.load(c_x_x)
print str(time.time()-t0)+' seconds to load dictionary.'

# CPI data parse

print; print 'Parsing CPI data.'
t0=time.time()

f=open('C:/Project/data/us_cpi_index/us_cpi_monthly.txt','r')
cpi_index_monthly={}
for l in f:
    cpidata=l.strip().split()
    cpi_index_monthly[cpidata[0]+'_'+cpidata[1]]=eval(cpidata[2])
f.close()

cpi_index_quarterly={}
for m in cpi_index_monthly.keys():
    if m.split('_')[1]=='1':
        cpi_index_quarterly[m.split('_')[0]+'_1']=cpi_index_monthly[m]
    elif m.split('_')[1]=='4':
        cpi_index_quarterly[m.split('_')[0]+'_2']=cpi_index_monthly[m]
```

```
    elif m.split('_')[1]=='7':
        cpi_index_quarterly[m.split('_')[0]+'_3']=cpi_index_monthly[m]
    elif m.split('_')[1]=='10':
        cpi_index_quarterly[m.split('_')[0]+'_4']=cpi_index_monthly[m]
    else:
        pass

# CPI deflator
# Hard-coded: 2009Q1 prices

cpi_deflator={}
for m in cpi_index_quarterly.keys():
    cpi_deflator[m]=cpi_index_quarterly['2009_1']/cpi_index_quarterly[m]

print str(time.time()-t0)+' seconds to parse CPI data.'

length=len(dt[dt.keys()[0]])

#print; print 'Length of data file: '+str(length); print
#print 'Dictionary keys: ',dt.keys(); print

print; print 'Creating output file.'
t0=time.time()

routes_all={}; carriers_all={}; timeslot={}
```

```
for i in range(length):
    if dt['OpCarrier'][i]!='TW':
        routes_all[dt['Origin'][i]+'_'+dt['Dest'][i]]=0
        carriers_all[dt['OpCarrier'][i]]=0
        timeslot[eval(dt['Year'][i]+dt['Quarter'][i])]=0

routes_all_list=routes_all.keys()
routes_all_list.sort()

route_codes={}; count=1
for j in routes_all_list:
    route_codes[j]=count
    count+=1

carriers_all_list=carriers_all.keys()
carriers_all_list.sort()

carrier_codes={}; count=1
for j in carriers_all_list:
    carrier_codes[j]=count
    count+=1

#print carrier_codes

timet=timeslot.keys()
```

```
timet.sort()
timecode={}; count1=1
for k in timet:
    timecode[k]=count1
    count1+=1

output='Origin'+'\t'+'Dest'+'\t'+'Year'+'\t'+'Quarter'+'\t'+
'OpCarrier'+'\t'+'AvFare'+'\t'+'TotalPass'+'\t'+\
 'Route'+'\t'+'Routecode'+'\t'+'Carriercode'+'\t'+'Time'+'\t'
 +'CPIquarter'+'\t'+'CPIdeflator'+'\t'+'RealAvFare'+'\n'

for i in range(length):
    if dt['OpCarrier'][i]!='TW':
        output+=dt['Origin'][i]+'\t'+dt['Dest'][i]+'\t'+dt['Year'][i]+
        '\t'+dt['Quarter'][i]+'\t'+dt['OpCarrier'][i]+\
         '\t'+str(dt['AvFare'][i])+'\t'+str(dt['TotalPass'][i])+'\t'+
         dt['Origin'][i]+'_'+dt['Dest'][i]+\
         '\t'+str(route_codes[dt['Origin'][i]+'_'+dt['Dest'][i]])+
         '\t'+str(carrier_codes[dt['OpCarrier'][i]])+\
         '\t'+str(timecode[eval(dt['Year'][i]+dt['Quarter'][i])])+'\t'+\
         str(cpi_index_quarterly[dt['Year'][i]+'_'+dt['Quarter'][i]])+'\t'+\
         str(cpi_deflator[dt['Year'][i]+'_'+dt['Quarter'][i]])+'\t'+\
         str(dt['AvFare'][i]*cpi_deflator[dt['Year'][i]+'_'+
         dt['Quarter'][i]])+'\n'
```

```
print str(time.time()-t0)+' seconds to create output.'
print; print 'Saving file.'
t0=time.time()

f=open(data_path+filename+'_text.txt','w'); f.write(output); f.close()

print str(time.time()-t0)+' seconds to save file.'; print
print 'Done.'

-----------------------------

"""

Data parse for dynamic economic fare models, by route.
(for project with Anthony/Maximilian)

Steve Lawford : 24 January 2010

"""

import os,sys
from scipy import *
from airport_location_function_modified_return import *
from route_population_geom_mean_function import *
from route_leisure_function import leisure_function
```

```
from route_business_function import business_function
from airport_gdp_function import gdp_function

module_dir=os.path.join('Project','code')
sys.path.insert(0,module_dir)

leisure_dict=leisure_function(0.05)
business_dict=business_function(0.10)
gdp_dict=gdp_function()

r_sphere=6372.7974775959065 # km

def distance_calculator(dest_coords_degrees,origin_coords_degrees):

    dest_coords_radians=[(pi/180)*dest_coords_degrees[0],\
                         (pi/180)*dest_coords_degrees[1]]
    origin_coords_radians=[(pi/180)*origin_coords_degrees[0],\
                           (pi/180)*origin_coords_degrees[1]]
    p1,l1=dest_coords_radians[0],dest_coords_radians[1]
    p2,l2=origin_coords_radians[0],origin_coords_radians[1]
    l=abs(l1-l2)
    num=sqrt(((cos(p2)*sin(l))**2)+(((cos(p1)*sin(p2))-(sin(p1)*cos(p2)
    *cos(l)))**2))
    den=sin(p1)*sin(p2)+cos(p1)*cos(p2)*cos(l)
    theta=arctan(num/den)
```

```
    distance=r_sphere*theta

    return distance

for rr in range(1,8):
    filename='fares_route'+str(rr)+'.txt'
    filenameout=filename.split('.')[0]+'_out.txt'
    f=open('C:/Project/data/'+filename,'r')
    c=0
    for j in f:
        c+=1
        k=j.split('\t')
        if c==1:
            lencheck=len(k); dct={}
            for i in range(len(k)):
                if k[i]!='' and k[i]!='\n': dct[i]=[k[i].strip(),['']*65]
        else:
            for i in range(len(k)):
                if k[i]!='' and k[i]!='\n': dct[i][1][c-2]=eval(k[i].strip())
    f.close()

    dct2={}
    for m in dct.keys():
        dct2[dct[m][0]]=dct[m][1]
    dct3={}
```

```
print
for n in dct2.keys():
    p=n.split('_')

    flag=0

    try:
        dest_coords_degrees=airport_location(p[0])[1]
        origin_coords_degrees=airport_location(p[1])[1]
        distance=distance_calculator(dest_coords_degrees,origin_coords_degrees)
        distance=abs(distance)
    except TypeError: flag=1

    try:
        popn=route_population(p[0],p[1])
        leisure=max(leisure_dict[p[0]],leisure_dict[p[1]])
        business=max(business_dict[p[0]],business_dict[p[1]])
        gdp=(gdp_dict[p[0]]*gdp_dict[p[1]])**0.5
    except KeyError: flag=1

    if not flag and popn!=None:
        dct3[n]=[dct2[n],distance]
        dct3[n].append(eval(popn))
        dct3[n].append(leisure)
```

```
        dct3[n].append(business)
        dct3[n].append(gdp)
    else: pass

output=''; lines=0
for s in dct3.keys():
    c=0
    for t in dct3[s][0]:
        c+=1
        if t!='':
            lines+=1
            output+=str(t)+'\t'+str(c)+'\t'+str(dct3[s][1])+'\t'+
            str(dct3[s][2])+'\t'
            output+=str(dct3[s][3])+'\t'+str(dct3[s][4])+'\t'+
            str(dct3[s][5])+'\t'
            if c>=35: output+=str(1)+'\n'
            elif c<35: output+=str(0)+'\n'

print str(lines)+' lines in file '+filenameout
f=open(filenameout,'w'); f.write(output); f.close()
```

List of Figures

List of Tables