Dynamic Networks:

with Application to U.S. Domestic Airlines*

Matthieu Dupont and Erwin Lodder

(supervised by Steve Lawford and Nathalie Lenoir)

DEVI, ENAC

January 27, 2017

Abstract

We investigate the dynamic properties of airline networks using tools from graph theory, specifically unweighted centrality measures. This provides insight into changes in network strategy over time, and into possible interdependence between the networks of different carriers. We also propose a simple time series forecasting model of airport centrality.

*We are grateful to Steve Lawford for supporting this project and to Nathalie Lenoir for having contributed her knowledge. This report was completed during a Mineure PIR – ECO project at ENAC, from September 2016 to January 2017.

1 Introduction

Networks play a central role in the strategic behaviour of airline companies. Carriers can change their network structure by opening (new) or closing (existing) routes between airports. In earlier work, Alexandre Dossin and Steve Lawford [1] show that local network structure, as measured by weighted and unweighted centrality measures, has a small but significant impact on the fares charged for direct flights. For illustration, they use static regression models, applied to the last quarter of 2013.

In this paper, we go further and examine how network structure and carrier strategy evolve dynamically. In particular, (a) we study how the estimated coefficients in fare regressions depend upon the time period used in the estimation, (b) we analyze unweighted centrality measures in order to identify salient features of domestic U.S. airline networks, and (c) we develop dynamic forecasting models for the time series evolution of centrality measures, that may provide insight into changes in airline strategy or interdependence between different airline networks.

2 Data and Model

Our main source of data is the U.S. Department of Transportation's Airline Origin and Destination Survey (DB1B), covering the period 1999Q1 to 2013Q4. The DB1B is a publicly-available 10% quarterly random sample of U.S. domestic airline tickets that is collected from reporting carriers. It contains information that includes operating and ticketing carriers, origin and destination airports, type of ticket, number of passengers, and total fare. We keep nonstop round-trip tickets between airports in the continental U.S., not including Alaska, that were not sold under a codesharing agreement.

Nominal dollar fares are converted to 2013Q4 prices, using U.S. Department of Labor price indexes. Individual tickets are then collapsed to a nondirectional route-carrier-quarter level. We discard tickets that have a nominal fare that is very low (below $20) or very high (above the 99th percentile of the route-carrier-quarter fare distribution). We also remove route-carrier quarters with fewer than 100 passengers in the DB1B. If more than 75% of a carrier's tickets are reported as business or first class, across all routes in

a given quarter, then we retain all tickets for that carrier, regardless of the fare class; otherwise, we retain only coach class tickets.

The DB1B data is extended with route-carrier-quarter characteristics from the U.S. Department of Transportation's T-100 Domestic Segment (All Carriers) dataset. This is a monthly 100% census of traffic data on domestic nonstop flight segments. We keep all passenger and cargo flights with more than 2,000 enplaned passengers or available seats at the directional route-carrier-quarter level.

Finally, the DB1B and T-100 data are merged to create the core dataset: U.S. domestic, nonstop, round-trip, coach class airline tickets, with no codesharing, over the years 1999 to 2013.[1]

2.1 Regression

In our static model, we regress the mean real fare of a carrier-route on carrier fixed effects, network variables and other explanatory control variables:

$$p_{ij} = \alpha + b_i + x'_{ij}\beta_{network} + w'_{ij}\beta_{controls} + u_{ij}, \tag{1}$$

where p_{ij} is the mean real fare for carrier i on route j; b_i is a carrier fixed effect; x_{ij} are network variables; w_{ij} are control variables; and u_{ij} is an error term. The method used for estimation is weighted least squares: since the mean fare is $p_{ij} = N_{ij}^{-1}\sum_{k=1}^{N_{ij}} p_{ijk}$, where k is an individual ticket and N_{ij} is the carrier-route pax, we can use weight $N_{ij}^{1/2}$.[2]

3 Econometric Results

We begin by analyzing the results of estimating model (1) on a single quarter: 2013Q4. In a second step, we automate the estimation to cover the full sample: 1999Q1 to 2013Q4. The aim is to analyze the impact on airline fares of each of the variables considered in model (1), and whether this changes over time.

[1] We do not use one-way or connecting tickets, and future work could consider these.

[2] The intuition for this is as follows: the mean carrier-route fare is based upon N_{ij} individual tickets, and so a higher value of N_{ij} will increase the precision of the mean fare; the weight corrects for the non-constant precision in the mean fare across the carrier-route observations in the sample, and enables us to use (weighted) least squares estimation correctly.

3.1 Static Results: 2013Q4

In the fourth quarter of 2013, there were 12 carriers present in the U.S domestic market. Five of these carriers are *legacy* airlines: American Airlines (AA), Alaska Airlines (AS), Delta Airlines (DL), United Airlines (UA), and US Airways (US). The seven remaining carriers are *low-cost* airlines: JetBlue Airways (B6), Frontier Airlines (F9), AirTran Airways (FL), Spirit Airlines (NK), Sun Country Airlines (SY), Virgin America (VX), and Southwest Airlines (WN). In the regressions that use carrier dummy variables, Southwest Airlines is the omitted category.

Visually, there seem to be some differences in the network strategies of these airlines (e.g. Figures 1 and 2). The low-cost carrier Spirit Airlines appears to have constructed a point-to-point network. Legacy carriers, on the other hand, often develop their networks as stars, with one or several very dominant hubs and more destinations served [2]. Moreover, legacy carriers typically serve the main airports in big cities (e.g. American Airlines at New York JFK) while low-costs tend to serve secondary airports, most of the time in the peripheral zones of the big cities (e.g. Spirit Airlines at New York LaGuardia LGA), or the main airports in smaller cities; see the report by Jérôme Morancé and Timo Noortman [3] for further details.

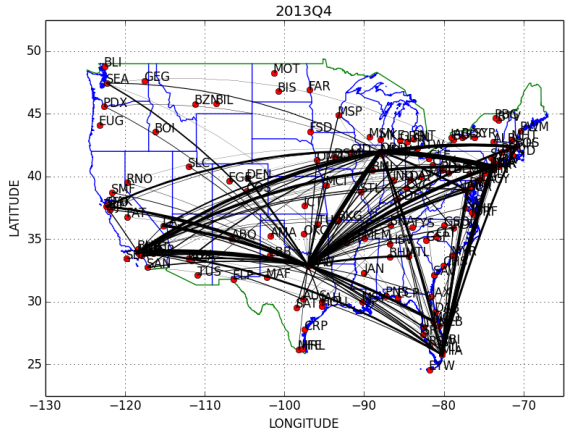

Figure 1: *American Airlines' (AA) network.*

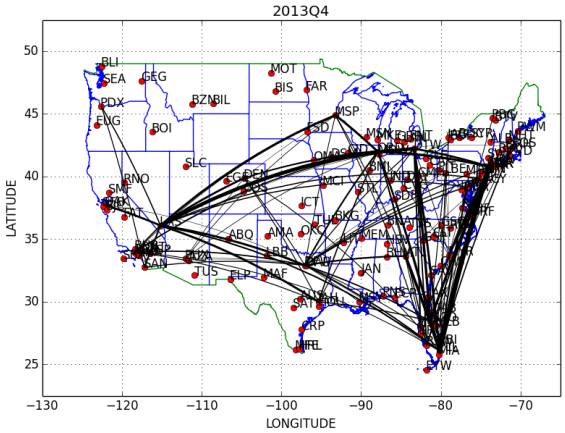

Figure 2: *Spirit Airlines' (NK) network.*

Do the differences in the network strategies of carriers have an impact on the fares that carriers charge? We attempt to answer this question by estimating model (1): the regression results are presented in Figure

A.1 in the Appendix. We include the following explanatory variables:

- *minbetweenness*: the minimum betweenness centrality across the endpoint airports of the route;
- *maxbetweenness*: the maximum betweenness centrality across the endpoint airports of the route;
- *distance*: the length of the route, in miles;
- *abstempdiff*: the absolute value of the temperature differential between the region of "departure" and the region of "arrival";
- *meangdppercapita*: the geometric mean of the Gross Domestic Product per capita in a 50 mile neighbourhood of the "departure" airport and a 50 mile neighbourhood of the "arrival" airport;
- $t100seats$: the number of seats offered by a given carrier on the route;
- *monopoly*: the route considered is a monopoly for a given carrier when the carrier serves at least 90% of the total number of passengers who travelled on the route;
- *competitive*: a competitive route is a route which is neither a monopoly route nor a duopoly route (on a duopoly route, the two largest carriers together serve at least 90% of the total number of passengers who travelled on the route);
- *WNpresence*: a dummy variable which indicates whether Southwest operates on the route.

Our model fits the data quite well insofar as the adjusted R^2 is equal to 0.75. If we take a look at the estimated coefficients of the different variables, some of these immediately confirm our intuition. For instance, *distance* has a positive marginal effect: longer flight distance results in a higher fare. The estimated marginal effect on route-carrier dollar fare is $20.32 - 0.36\,(distance/100)$, where *distance* is measured in miles. A 100 mile increase in route length increases fares by \$18.51 for a (short) 500 mile route, and by 13.09 for a (long) 2,000 mile route. The main explanation for this is the higher cost of fuel, which is one of the largest direct costs for an airline.

A monopoly route also has a positive marginal effect on fares (+\$32.06 compared to a duopoly route), while a competitive route has a negative marginal effect (−\$18.96 compared to a duopoly route). This shows that airline markets respect the textbook competitive market laws: if an airline is the only operator of a route, it can take advantage of this situation by setting higher prices for tickets.

However, we note some curious facts, including a negative marginal effect for the legacy carrier Alaska Airlines (AS) compared to the omitted category, the low-cost carrier Southwest (−\$39.88); and a positive marginal effect for Virgin America (VX), which is considered to be a low-cost (+\$28.05 compared to Southwest, on routes on which Southwest *is not* present, and −\$54.87 compared to Southwest, on routes on which Southwest *is* present).

The last important thing we notice is the significant impact of the local structure of the airline network, as shown by the positive marginal effects of the centrality measures: the variables *minbetweenness* and *maxbetweenness*, in this case. A 0.1 increase in the minimum (maximum) betweenness centrality at an endpoint airport — controlling for carrier, route length, route competitive structure, etc. — corresponds to a +\$23.78 (+\$2.81) change in mean fares. Hence, when the least central endpoint on a route becomes more central (in this case, when the airport is on 10% more of the shortest topological paths between any pair of airports in the network), the impact is ten times greater than if the most central airport becomes more central. This suggests that the more central is an airport for an airline, the higher will be the price of tickets on the routes to and from these airports.

Large network carriers make passengers fly through their hubs and seem to fix higher prices on these routes. Indeed, these carriers dominate operations at one or more hub cities. With domination comes a degree of market power and attractiveness to passengers, enabling the carrier to extract a fare premium on flights to and from the hub [4]. Additionally, domination provides a means to exclude new competition through control of facilities and aggressively competitive or even predatory practices [5].

3.2 Dynamic Results: Evolution of Coefficients

We wrote a Python code to investigate the dynamics of model (1): the idea is to examine the impact of each explanatory variable through time rather than for a single quarter. Our code first estimates the mean real fare model for a given quarter. It then stores the coefficients of each variable for that quarter, before looping over all quarters in the sample. Finally, it plots the time series of each estimated coefficient. These dynamics allows us to identify some trends in the impact of the different variables on fares. It also gives us some ideas about possible changes in the behaviour of the carriers, the impact of some local or global events on the airline market, and the evolution of this market.

In this part, we will first study the evolution of the impact of monopoly routes on airline fares (see Figure 3): this will give us some indication about changes in the way in which carriers are able to translate market power into higher fares. Then, we we will focus on the behaviour of a single airline, American Airlines (see Figure 4).

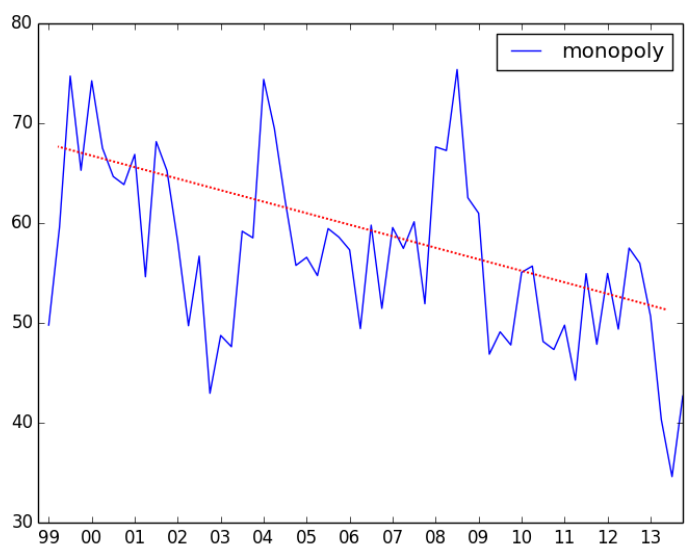

Figure 3: *Estimated monopoly coefficient from model (1), for 1999Q1 to 2013Q4.*

On a monopoly route, a single carrier has more than 90 percent of the total number of passengers on that route. The coefficient, as expected, is always positive: airlines are able to take advantage of

monopoly routes by setting higher prices on these routes than on routes where there is greater competition. Interestingly, the coefficient falls linearly over time, suggesting a reduction in the impact of market power during recent years. If we look at the airline market, we find two possible explanations for this situation:

- First, two carriers could each have a monopoly on two neighbouring but different routes. For example, a legacy carrier could have a monopoly on a route from airport A to B and a low-cost carrier from A to C, where A and C are geographically close to one another.
- Secondly, with the development of low-cost carriers, connecting tickets between A and B, with a stop at D, are sometimes cheaper than a direct flight from A to B with a legacy carrier that has a monopoly on the route.

In both situations, the carriers offer similar service to a direct trip from A to B: non-direct transportation from the same origin to the same destination, or direct service from the same origin to a nearby destination. The passenger has multiple options for the same service, so the market power goes down. (Of course, this assumes that other ticket characteristics are similar e.g. fares, schedule, restrictions).

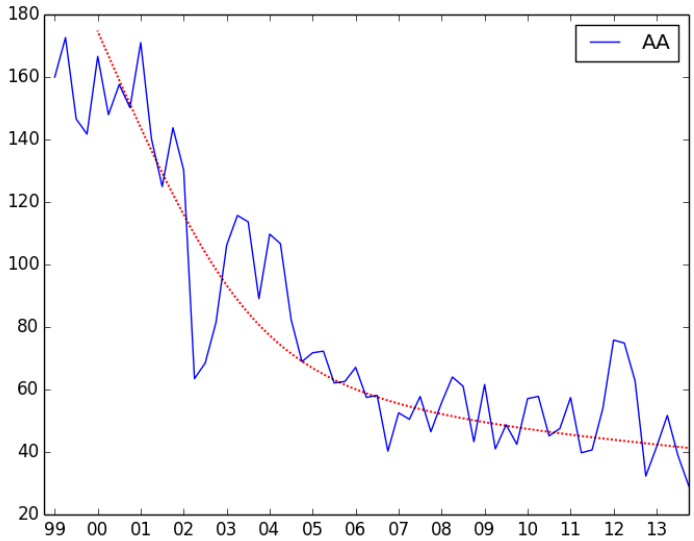

Figure 4: *Estimated American Airlines indicator coefficient from model (1), for 1999Q1 to 2013Q4.*

The coefficient of American Airlines, relative to Southwest, falls quadratically and is divided by four between 1999 and 2013. This shows the tendency towards convergence of fares between legacy and low-cost carriers. Indeed, during the last few years, the low-cost carriers have developed flight connections. This increases their operating cost and the price of their tickets. Meanwhile, legacy carriers want to remain competitive so they significantly decrease their profit margins.

Then, we can identify the impact of events such as the terrorist attack of September 11, 2011. This global event caused a significant decrease in passenger demand in the months following the attack. It resulted in a fall in airlines tickets fares, which was more significant for legacy carriers than for low-cost carriers insofar as legacy margins were previously higher and easier to reduce.

The decrease between 2004 and 2005 corresponds to a change of CEO, with the nomination of Gerard Arpey. It is hard to confirm but we can imagine a new strategy adopted by this new CEO which was to reduce its margin in order to become more competitive.

4 Understanding Centrality Measures

In order to model the mean real fare of a carrier on a route using forecasted network properties, we can try to model the centrality measures. This would result in a fare tool, which is more dependent on changes in the airline's network. If an airline decides to cancel a route, thereby modifying its network, we should observe an impact on the aire fares on remaining routes. Centrality measures have not often been modelled before, and then not for airline networks: Hochberg, Ljungqvist and Lu [6] modelled centrality measures for a first time venture capital firm's network position, using IPO transactions and investments in year $t-1$. Comparisons will be made at the airport, area and country level, equivalent to single airports, airports in urban areas and U.S. regions, respectively.

4.1 Nodes

Four centrality measures are given for each airport: degree, betweenness, eigenvector and closeness centrality. Not all of these are equally worth investigating: the interpretation of eigenvector and closeness centrality is much less clear. Degree and betweenness centrality both refer to the importance of an airport's position in an airline's network, but not in the same manner. Although degree and betweenness are strongly correlated, they can give very different results for an individual airport. A node can have fewer edges, but a high betweenness. A good example is Chicago Midway MDW in Southwest's network in 2013, which has a significantly higher betweenness while sharing a equal level of degree centrality with other airports. Figure 5 and 6 show the relative centrality levels of Southwest's airport in the U.S.. This is due to some destinations being connected only through Midway airport.

The variables conditioning airline activity and making airlines successful on routes arise from a combination of passenger numbers and air fares. Would there be a connection between these airline activity parameters and centrality measures, it would be present with betweenness or degree centrality. As an example, Figure 7 shows the betweenness measure of John F. Kennedy for JetBlue, enabling the opportunity of comparing both constants over a 60 quarter time span. The second axis displayed on the

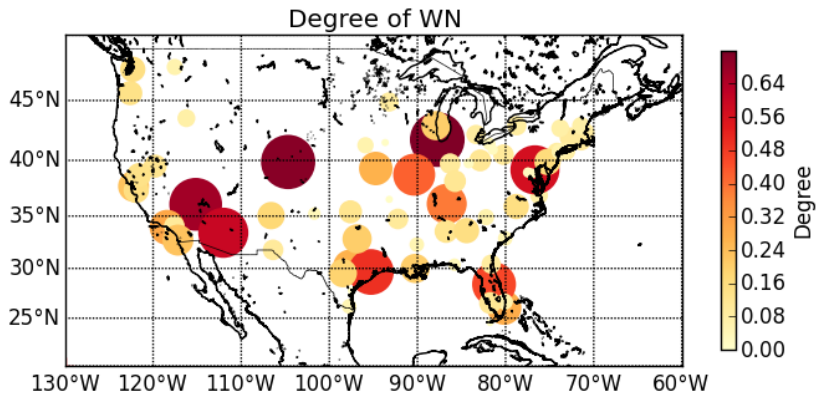

Figure 5: *Degree centrality for Southwest in 2013.*

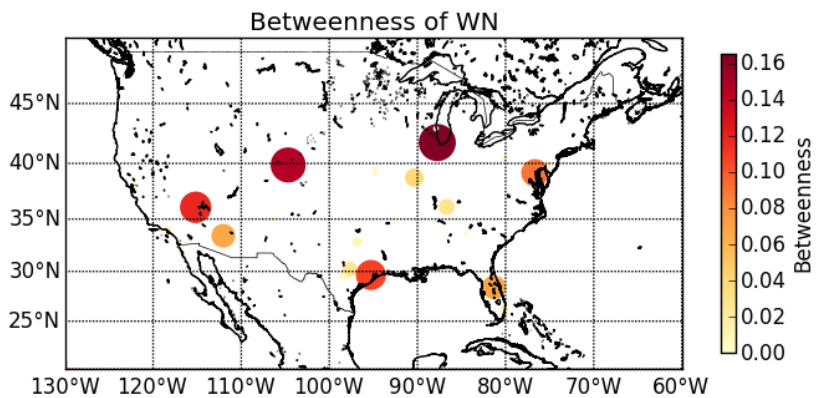

Figure 6: *Betweenness centrality for Southwest in 2013.*

right of the figure is the cents per mile (CPM) variable. The CPM is a measure that combines air fares with distance. This number tells us how much the passenger is paying per travelled mile. The average cents per mile for one quarter is calculated using (2). This adjusts the airfare with the proportional number of passengers that have travelled on a particular route i.

$$CPM = \frac{fare_1 \times pax_1 + fare_2 \times pax_2 + \cdots + fare_i \times pax_i}{distance_1 \times pax_1 + distance_2 \times pax_2 + \cdots + distance_i \times pax_i}. \tag{2}$$

For a realistic comparison it is not correct to use all routes from an airport for an average CPM. As the distance of a route increases, the key components forming a ticket price will change. For example, the fare for a short flight from San Francisco SFO to Los Angeles LAX will be made up relatively more by airport handling costs than a flight from SFO to Miami MIA. Also, differences in strategies apply to different

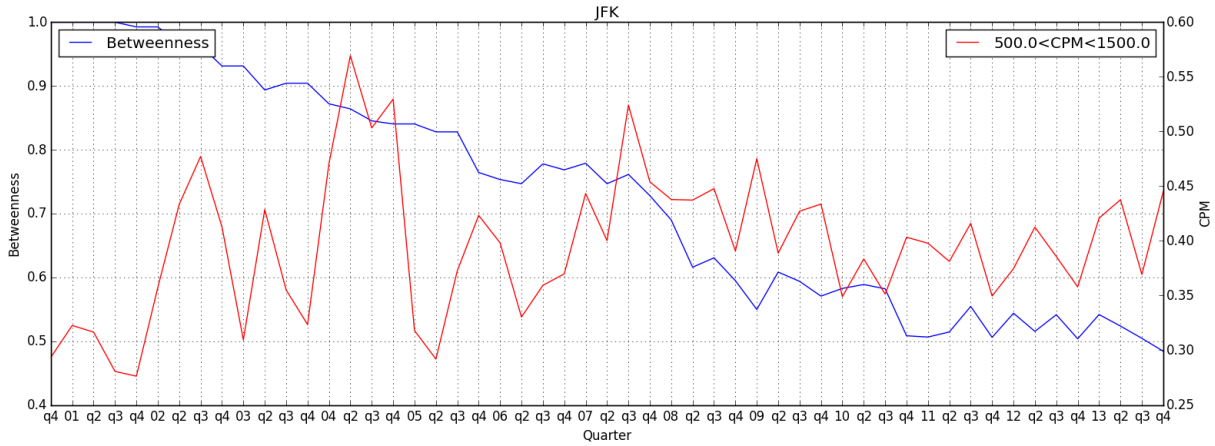

Figure 7: *JetBlue's betweenness and average CPM at John F. Kennedy airport.*

distances, meaning that revenue margins for airlines change on their routes as distance increases. Since strategies on short- and medium-haul routes differ for airlines, an average of airfares on all routes could give unreliable results. The distances are set at 0–500, 500–1500 and 1500–5000 miles, and it is assumed that all flights within these intervals belong to the same category.

Airports with newly introduced airlines over the sample time span might give an idea about the evolution of air fares and passengers for that particular airline or of its competitors. This applies equally to airports that experience the opposite. An airport that saw the arrival of a new airline in 2006 was Denver. As can been seen in Figure 8, Southwest began service at this airport from 2006Q1, making it a base in 2006Q2.

All three figures show inconsistent CPMs until 2010, gradually increasing afterwards. However the distance category with the most flights, for example 27 routes between 500 and 1,500 miles in 2013, shows a steady increase from 2006. This can be explained by a strategy where Southwest introduces low fares in order to gain more passengers when starting operations. By comparing the developments of other airlines at Denver a more realistic view is given. For convenience, all airlines operating from Denver that have more than 10 routes are shown. The next section will elaborate on determining the minimum number of routes. Figure A.6 in the Appendix shows the huge passenger increase of Southwest in combination of low cost carrier Frontier, which might have caused the sudden departure of United Airlines. United Airlines has been expanding in Denver; however, this changed with the arrival of Southwest. From 2006Q2

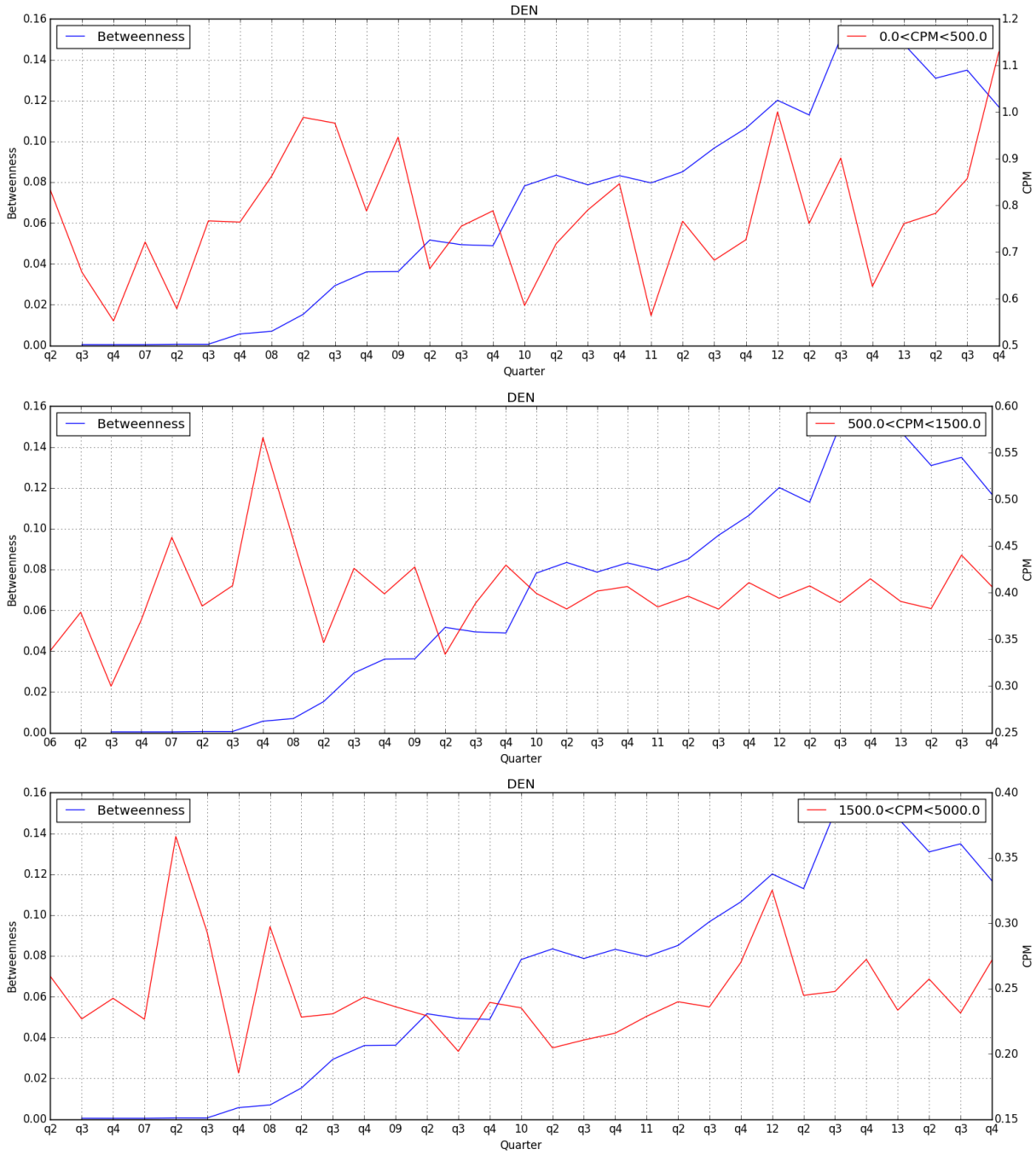

Figure 8: *Southwest's betweenness centrality plotted against different CPM's (Denver).*

onwards, United's presence in Denver reduced, while Southwest strengthens its position. The decrease in United Airlines's betweenness centrality shows a strong link with the speed of the increasing market share of Southwest in Denver. This means that the speed of growth or decline in market share for airline

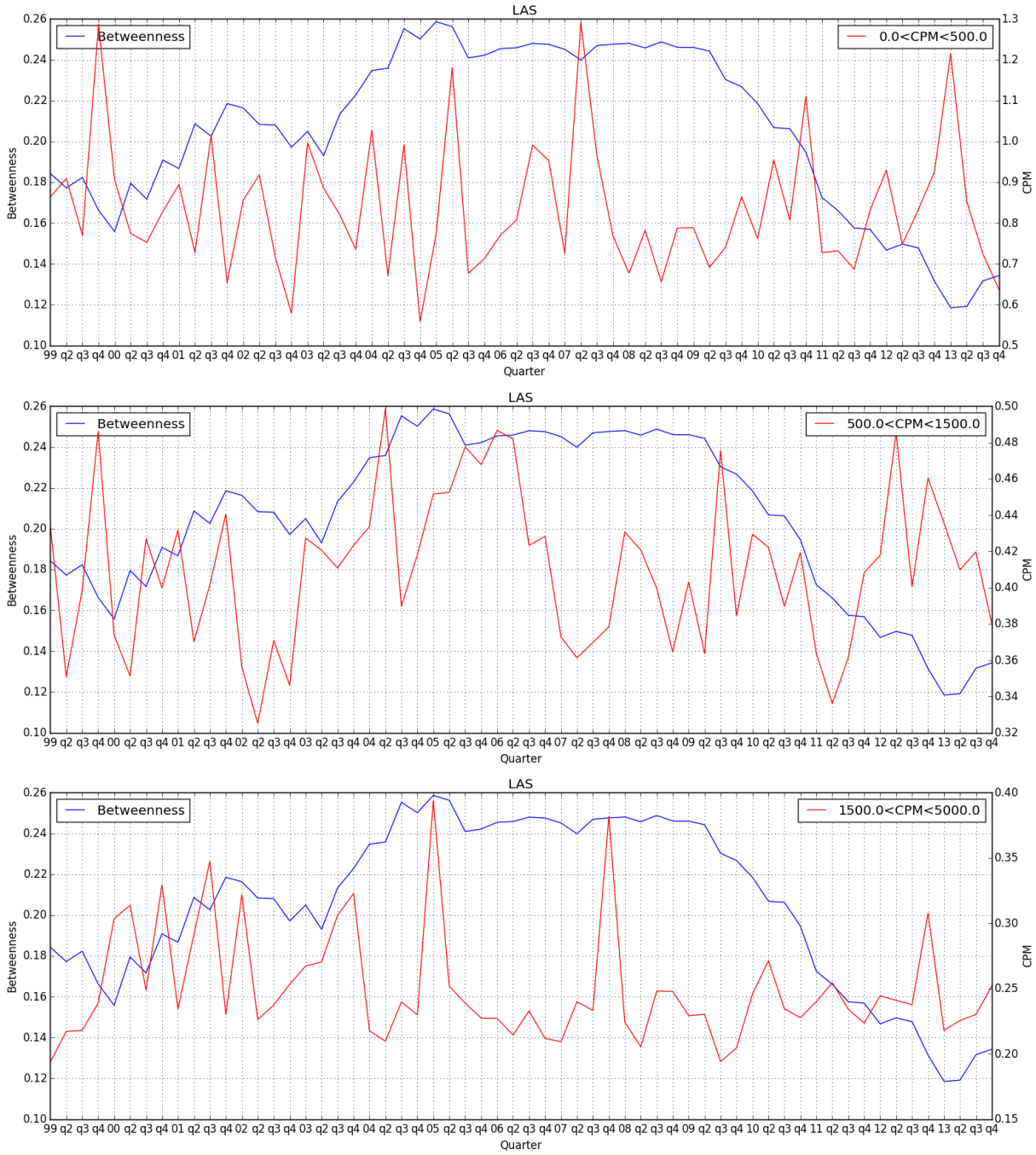

Figure 9: *Southwest's betweenness centrality plotted against different CPM's (Las Vegas).*

A affects the increase or decrease of airline B's centrality at the same airport. As Southwest has shown to be of great influence on an airport, other airports in Southwest's network with the same or reversed development in betweenness centrality are interesting to examine. Figure 6 shows other key airports in

Southwest's network, such as Las Vegas McCarran Airport. McCarran Airport is particularly interesting as it experienced a strong and weak hub position as can been seen in Figure 9. Except for the mile range of 500 to 1500 between 2003 and 2006, the trend shows that a higher betweenness goes with a lower CPM. At hubs large and small aircraft can be changed according to the capacity needed per route, keeping the operational costs lower and therefore declining the CPM [2].

At first the graphs in Figure 9 seem contrary to the explanation given before, but there is a difference between the airports analyzed. The difference in behavior can be caused by the occupancy degree of Southwest flights over all flights, which is much higher at McCarran airport than Denver. Especially between 2005 and 2009 Southwest dominated McCarran resulting in cheaper fares because of a size advantage. To confirm this assumption Figure 10 shows the same trend for Southwest at Chicago Midway airport, namely declining ticket fares while the betweenness is growing. This also occurs for legacy

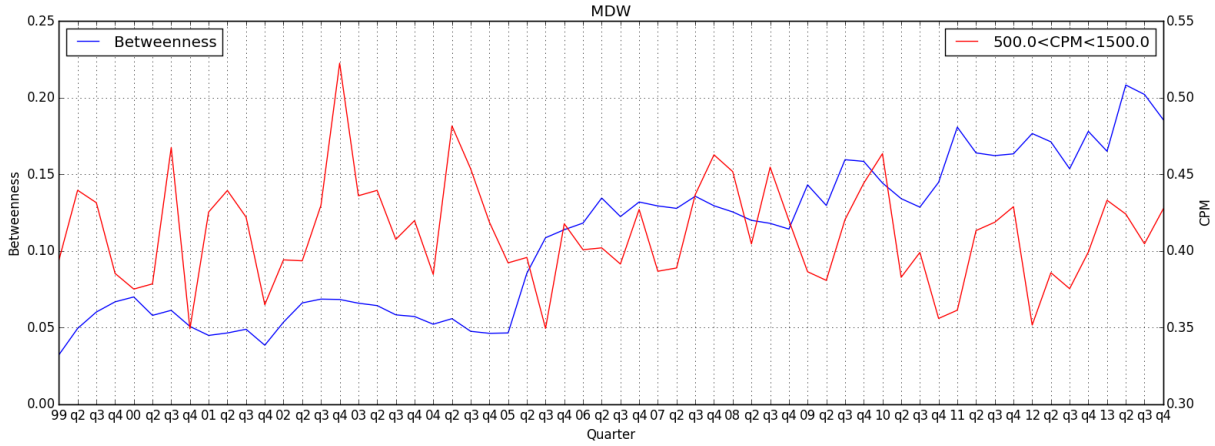

Figure 10: *Betweenness centrality plotted against CPM at Midway for Southwest.*

airlines such as American. Figure 11 again confirms the previously explained patterns.

For airports with a more evenly distributed market share over multiple airlines it is difficult to get an impression regarding fare and centrality causalities. As an example Los Angeles airport is given, which is used by six airlines as a base. Once again an adapted minimum number of flights is chosen for an airline calling Los Angeles a base or hub. Visually it would not be possible to examine these plots as they are too unclear. A closer look at Los Angeles and the surrounding airports might give more usable information as elaborated on in the next section. There are a lot of interesting configurations and situations between

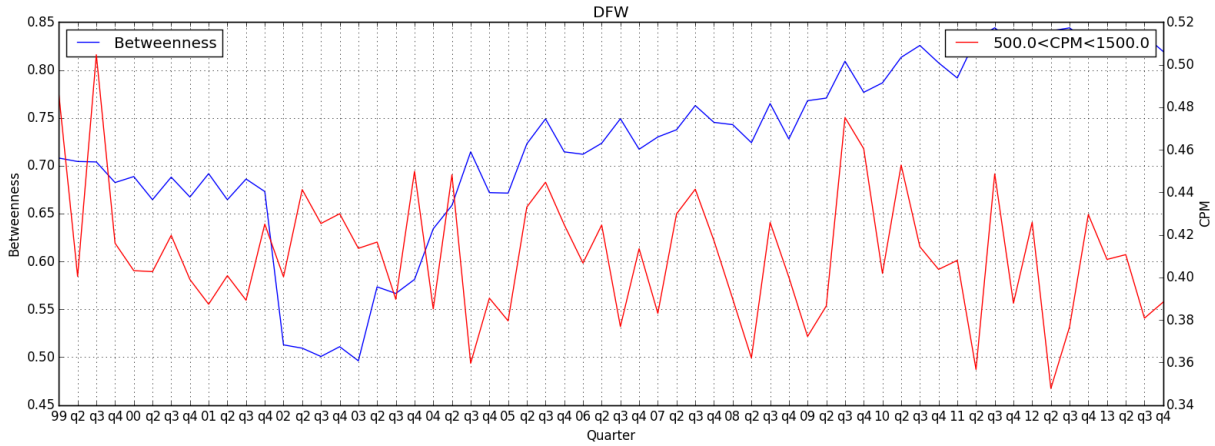

Figure 11: *American's increasing presence together with a decreasing air fare.*

airlines, market shares, centralities, airport and passengers but all require a different type of approach for investigation which is, because of time constraints, not discussed here.

4.2 Nodes in the Same District

Reviewing single airports only suffices to give interpretations for airlines at that airport. For cities with multiple airports, it is highly probable that these influence each others centrality in dense populated areas. Therefore it is necessary to evaluate neighbouring airports. A good example are the airports in Chicago: O'Hare and Midway. Both airports need to be examined simultaneously since airline activity at O'Hare might change according to the air fares or passengers at Midway. Figure 12 shows a strong correlation between the increasing presence of Southwest and the other legacy operators that only serve O'Hare. Although the betweenness centrality for Southwest is lower, the airline can have more connections out of Chicago than the legacy carriers. This is due the number of edges in Southwest's network, increasing the density and decreasing betweenness centrality of a particular node. Southwest is experiencing a steady growth at Midway while American Airlines and United Airlines are negatively impacted by a reduction in activity. The similarity with Denver is the increasing presence of Southwest while pushing out other legacy airlines. Nevertheless, the difference is the speed of expansion and decline of both operating parties. The small passenger growth of Southwest influences the mellow departing rate of United and American.

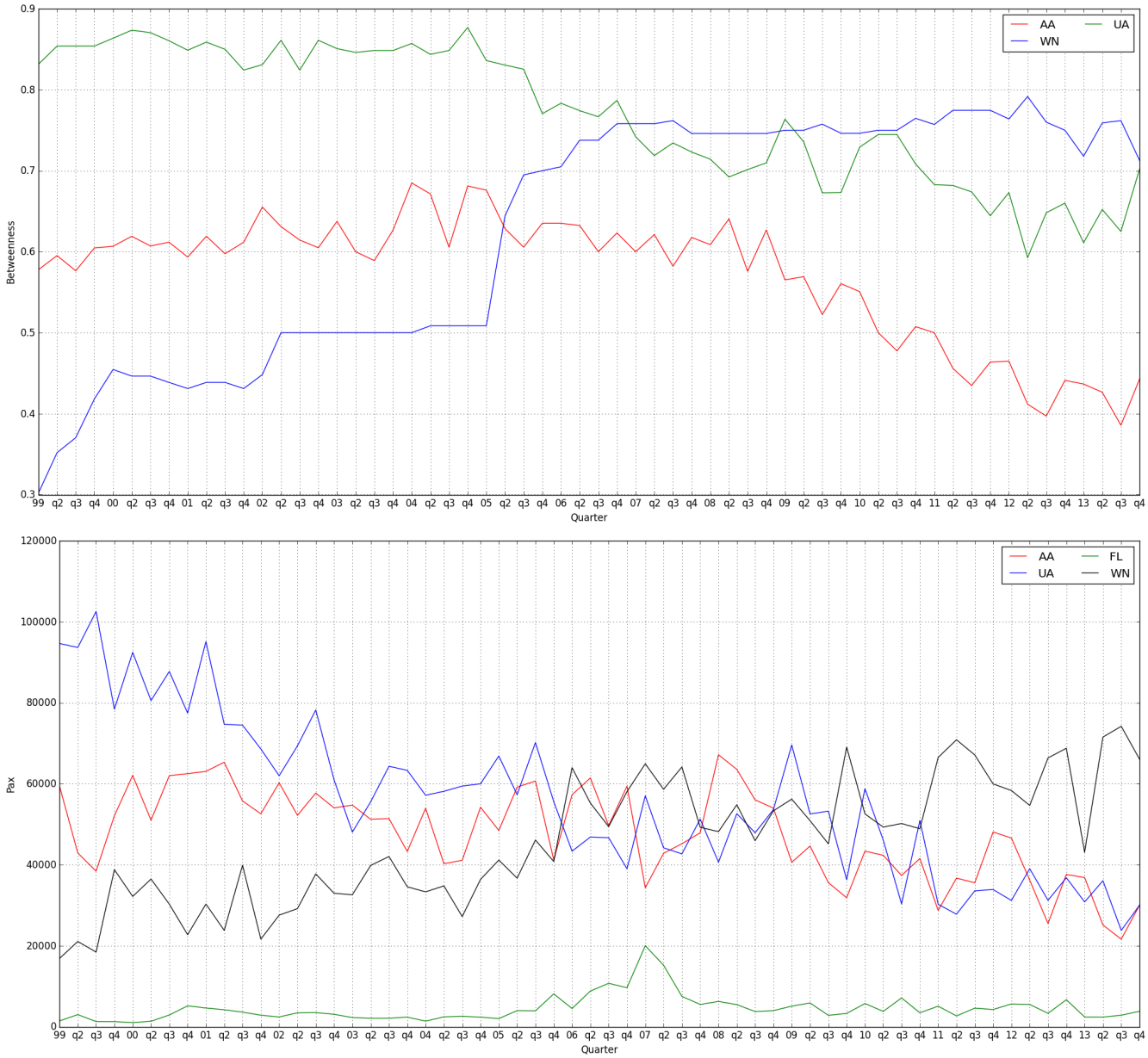

Figure 12: *Betweenness centrality and carried pax for airlines out of Chicago.*

In contrast to Denver where United has been leaving rapidly.

The CPMs of these airlines could also be plotted, but this would make the figure too difficult to interpret. While there are more airlines serving the Chicago airports, not all of them use one of these airports as a hub or key airport. This also means that although, for example, Delta serves Chicago from more than one origin, Delta has no significant contribution to the market in Chicago. Therefore we choose

to set the minimum number of connections to Midway or O'Hare to seven, filtering airlines with a less considerable market share. This is not completely fair as the less important airlines can have an impact on the fares and passengers from airports in Chicago, but it is less likely that they do.

The minimum number of flights from an airport needed to appoint an airline as important is hard to determine. Not only does the size (in terms of passengers) make a difference, the type and surrounding airports can change this minimum number of flights. During this research these values are approximated by analyzing every airport separately and adjusted if airlines with a significant low market share are present in the figure.

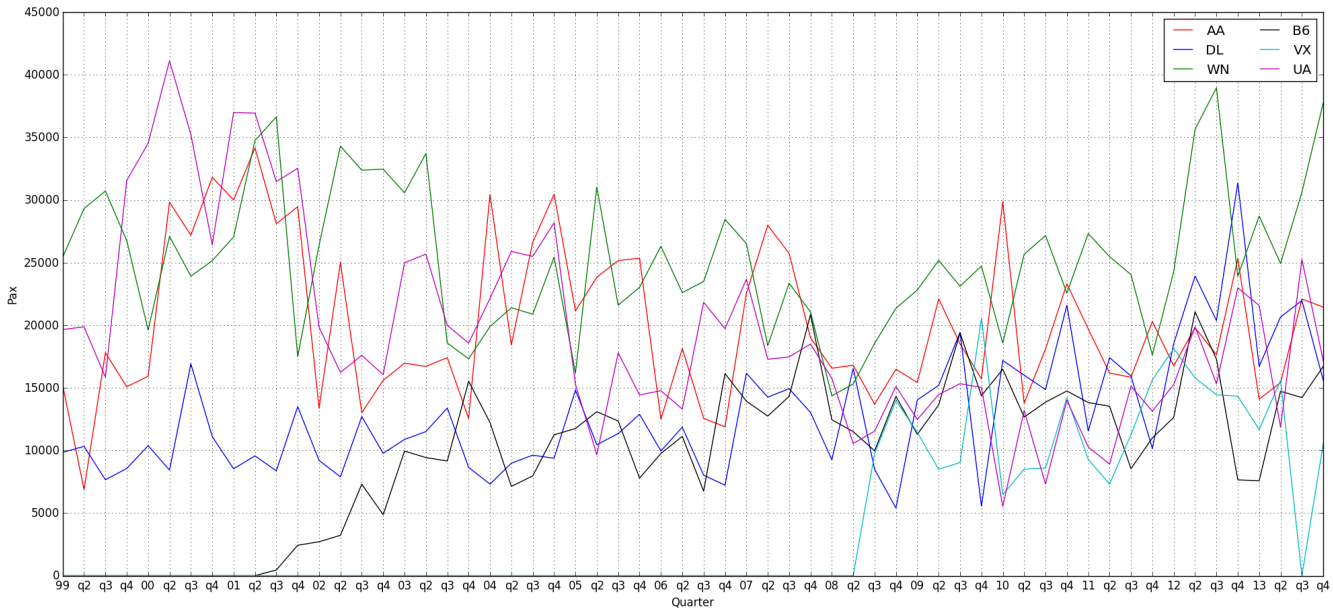

Figure 13: *Number of passengers divided by* 100 *in the urban area of Los Angeles.*

Another urban area with more airports is the area around Los Angeles, which counts five airports: Los Angeles, Ontario, Long Beach, Burbanks and Santa Ana John Wayne airport. The number of airlines representative in and around Los Angeles is high, without a clear hub position for any of them. Figure 13 shows the total passengers each is carrying out of the area of Los Angeles. As stated before it is hard to interpret such graphs. Although a mathematical analysis is possible, processing these values for further conclusions requires every urban area to examined separately. Calculating the effects of betweenness and degree is hard in regions like Los Angeles, because some airlines have sub hubs among neighbouring

airports. It is not correct to simply add up centralities as the sum is not the representative centrality of a region for an airline's network. Adding up passengers is however a valid parameter to use for urban areas. For future calculations a method needs to be invented assigning a parameter for an airline's position in a region as Los Angeles. This parameter would allow further analysis for a region instead of every airport separately.

4.3 Node Clusters

Before, it was observed that United and American have lost territory in Chicago. How did their strategy change? In order to understand the consequences caused by Southwest in Chicago, a broader view is needed. Figures 14 display a centrality distribution of a particular airline in one quarter. The radius and colour of the circles indicate the level of centrality. For United this results in pulling out of Chicago and taking in a more evenly-spread network. American starts focusing on a star network by moving flights to Dallas Forth Worth.

Figure 15 displays the degree centrality of the legacy carrier Delta and low-cost airline Southwest in the first quarter of 2006. The first impression is that Southwest operates significantly more destinations, as more airports are clearly distinct. Although Southwest operates more flights, the difference in destinations is not huge. For Delta, Atlanta is the primary hub, almost serving every other destination Delta has in their network. Southwest has a different approach, offering more connections over more airports resulting in a higher density and centrality at each airport.

These figures can give a good interpretation of developments of certain areas and airports. After analyzing all airlines for 60 quarters, it is evident that every 5 years or so there appears a new strategy. This does not apply for star networks as there is only one prominent airport. As example Southwest and American are analyzed for changing strategies. Figure 16 shows that in 2000 Southwest spread its flights, centralizing its activities in 2008 and at last in 2013 spreading a higher level of betweenness among more airports. American starts with a dual star shaped network in 1999, spreading its flights over 4 hubs and finally decides in 2013 to use Dallas Forth Worth as their key airport. Once a different strategy has been

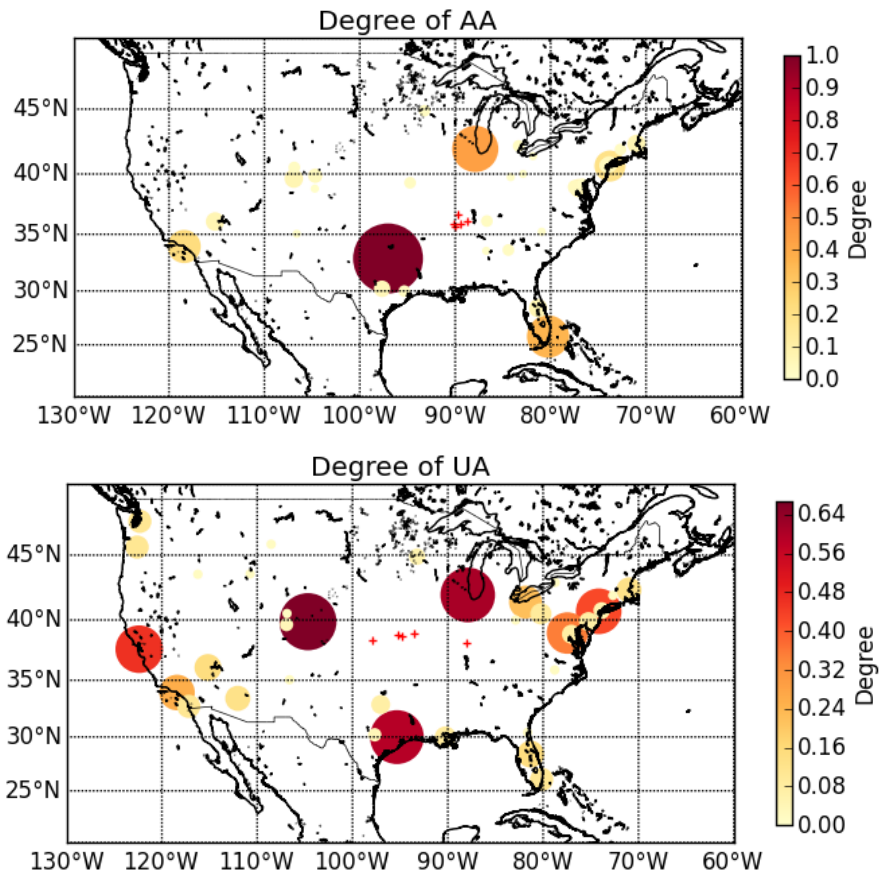

Figure 14: *Degree centrality distributions over the U.S. in 2013, while the red crosses show the shifting centre of gravity passengers and airport locations.*

adopted, it is less complicated to think about the resulting changes for centralities and other airlines at the influenced airports.

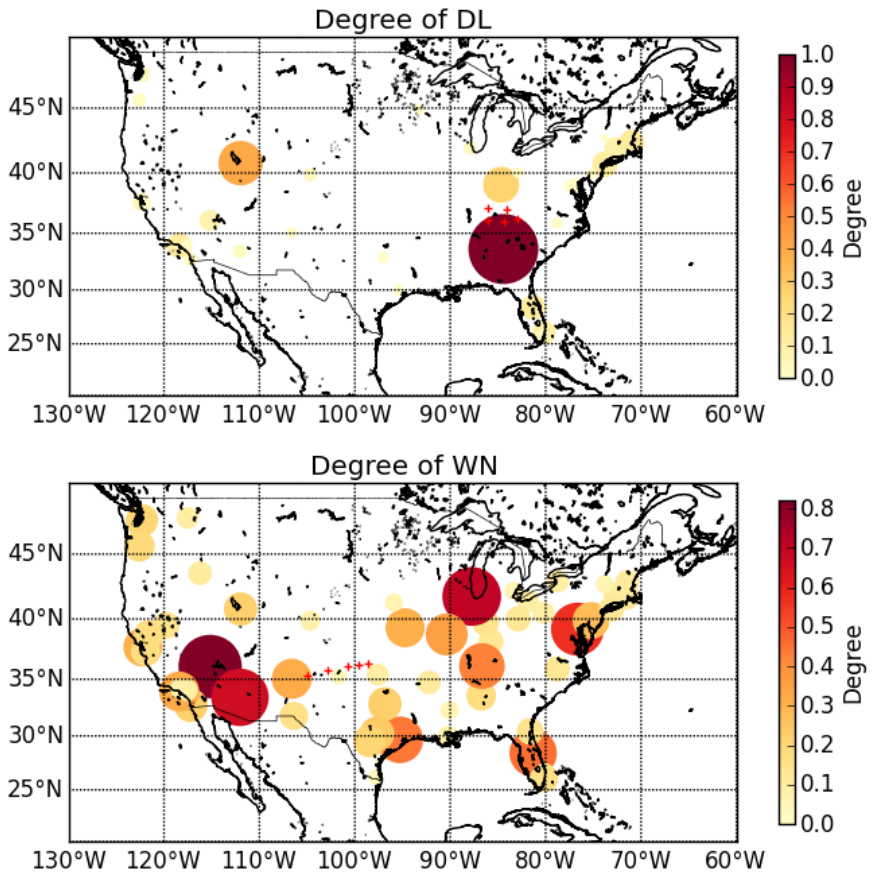

Figure 15: *Degree centrality distributions over the U.S. in 2006, while the red crosses show the shifting centre of gravity passengers and airport locations.*

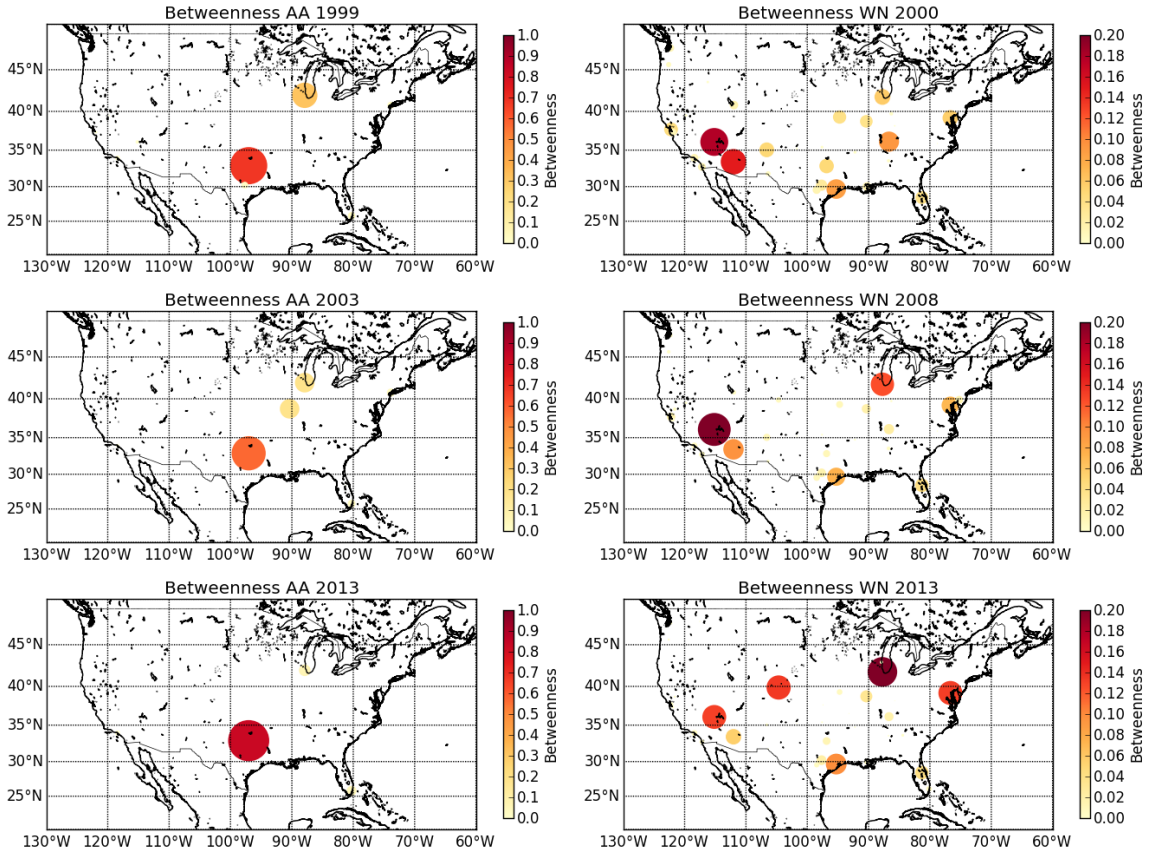

Figure 16: *Betweenness distributions for Southwest and American for different years.*

4.4 Conclusion

While trying to process and plot the data, a lot of constraints became apparent. The biggest issue was merging airlines, because they are unsuitable for research as a merger causes a completely new network. Subsidiaries were also not taken into account, while in many cases they carry out a meaningful part of flights. Mostly short haul operations are performed by subsidiary airlines, however carrying the livery of the airline flying for. Plotting fares also showed how powerful airline strategies are since a lot of graphs show sudden considerable changes which ruin observable trends in ticket prices. In contrast, centralities have been shown to be a more constant factor as airlines do not suddenly move a substantial amount of their operations to a new or existing airport in their network. Airline networks are unique in a sense that every node has very specific properties. Properties as the size of the airport, type of airport and peak times all change the relative value of centralities for airlines. Also the edges can not only be seen as a simple connection, since frequencies, departure times and types of flights (leisure or business) change the importance and value of a route.

5 Forecasting Centrality Measures

Now we have developed some intuition about the centrality measures of airline networks, the natural follow-up is to try to forecast these measures. Indeed, by the previous analysis of the centrality measures, we have seen that the network of an airline is in constant evolution and that some airports become more or less important in an airline network through time. The forecast of the centrality measures of the airports could enable us to predict some characteristic behaviour of an airline network and maybe some interdependence between several airlines networks.

The idea is to drop variables from the econometric baseline model, one at a time, in order to develop some intuition about the correlation between centrality measures and other variables. Then, we will try to build some models to forecast centrality measures.

5.1 Correlations with Centrality Measures

Figure 17 shows us on the left the value of the centrality measures (*minbetweenness* and *maxbetweenness* here) when we remove a single variable from our first econometric model. On the right, the value of the adjusted R^2 enables us to judge if the quality of the new model is much affected. The goal is to highlight some correlations between the centrality measures and some variables we have considered.

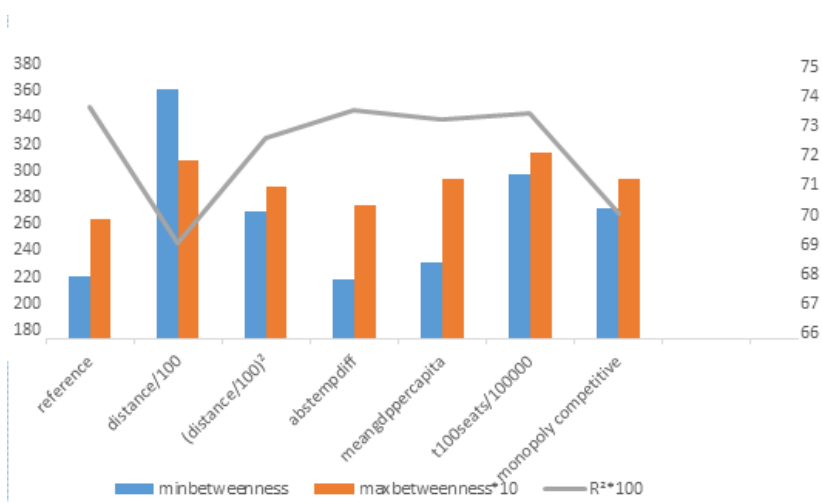

Figure 17: *Influence of a variable removal from our baseline econometric model.*

For example, when we drop $meangdppercapita$, the centrality measures become more important, suggesting a positive correlation between the centrality measures and $meangdppercapita$. Thus, the most important airports seem to be developped in the richest regions of the U.S. So we certainly have to take into account the wealth of the region of an airport when we will try to forecast its centrality measure.

Another example is the important positive correlation between distance and centrality measure. The further you want to go, the more likely you are to fly to or from an important airport. It is in compliance with legacy strategies which make you fly by their dominant hub when you want to cross the country. However, the quality of the model decreases by more than five percent so we will have to be very careful when we will use this information in our estimation.

5.2 Estimated Model

We will use the same data that we used in our econometric model for fares, but now at an airport-carrier-quarter level rather than at a route-carrier-quarter level. Now, the explanatory variables that we used have the following meanings :

- *betweenness*: the betweenness of the airport in the network of the carrier considered;
- $monopolyroutes$ ($competitiveroutes$): the number of monopoly (competitive) routes served by the carrier at the airport considered;
- *pax*: the number of pax carried by the airline on all routes out of the airport;
- *capacity*: the number of seats offered on all routes out of the airport by the given carrier;
- $loadfactor$: the total number of passengers on all flights out of the airport divided by the total capacity on all flights out of airport for the given carrier;
- $nodes$: the total amount of nodes in the network of the carrier considered;

We will fix a carrier and an airport to establish a model for the centrality measure of this airport in the network of the carrier considered. Insofar as we have a huge amount of airlines and airports, the point will

be to see if we can find some recurrent behaviour between the different networks evolutions of the airlines.

To test the accuracy of our forecast model, we will evaluate different models from 1999Q1 to 2011Q4 and run a dynamic forecast from 2012Q1 to 2013Q4. Because we have the real data for 2012Q1 to 2013Q4, it will enable us to see the gap between our forecast values and the real values of the centrality measures.

As we had to fix a carrier and an airport, we chose American Airlines at Miami airport. This airport is not a major hub for American Airlines concerning continental flights and a lot of airlines are present at this airport, so this airport is involved in a lot of airlines networks.

For both of our models, the centrality measure we estimated was *betweenness*. For our first estimation, we used the first lagged values of *betweenness*, *monopoly*, *competitive* and *pax* as explanatory variables. With this simple model, we have a positive influence of the lagged value of *betweenness* insofar as the expansion of an airport in an airline network follows a trend in the same direction: if an airline chooses to give influence to an airport in its network, it is unlikely to change its mind in the following quarter. Added to this, we expect a positive marginal effect of *monopolyroutes*: legacy airlines open monopoly routes from their dominant airports. It follows their general strategy which is to make passenger fly through their hubs before reaching their final destination. In a same way, we should have the same result for the *pax* variable. However, neither *monopolyroutes* nor *pax* are significant in this model. (see Figure A.2 for the table of the results)

The results of the dynamic forecast are quite far from the real values, but the variations are similar (curve given by Figure A.3). This is a good point if we only want to forecast the trend of an airport's influence in an airline network.

In the second model (figure A.4), we added the lagged variable of *capacity*, which is the airport-level capacity (number of seats) offered on all routes out of the airport Miami by the chosen carrier, American Airlines. As for *pax*, we expect that the more capacity an airline offers at a given airport, the more it will give this airport some importance in its network. This is the case for this model: if capacity rises by 100,000 seats, *betweenness* rises by 0.015 in the next quarter. This is not aberrant: an increase of 100,000 seats in a given quarter corresponds to 5 more flights of 150 passengers per day from this airport.

The variable *nodes* corresponds to the number of nodes in the whole network of American Airlines. Its

impact enables us to know if a change in the global American Airlines network affects the Miami airport locally. Here we can see that an expansion of American Airlines network in term of nodes will result in a decrease of the *betweenness* of Miami airport. So, if American Airlines chooses to serve more airports, the importance of Miami airport in its network will decrease in term of betweenness.

The quality of this new model rises by 8% comparing to the previous model. Concerning the forecast (figure A.5), the model works pretty well for the year after but the results are really inaccurate after one year.

Previous variables only include values extracted from the data set. As we saw a correlation between betweenness centrality and passengers, using passengers as weights and airport coordinates as locations to calculate the geographical centre of gravity (COG) looks promising. The COG gives us a good interpretation of the shifting edges over the network. Especially for airlines with one, two or three dominant hubs, a change in COG positions means that connections disappear and form. Figure 18

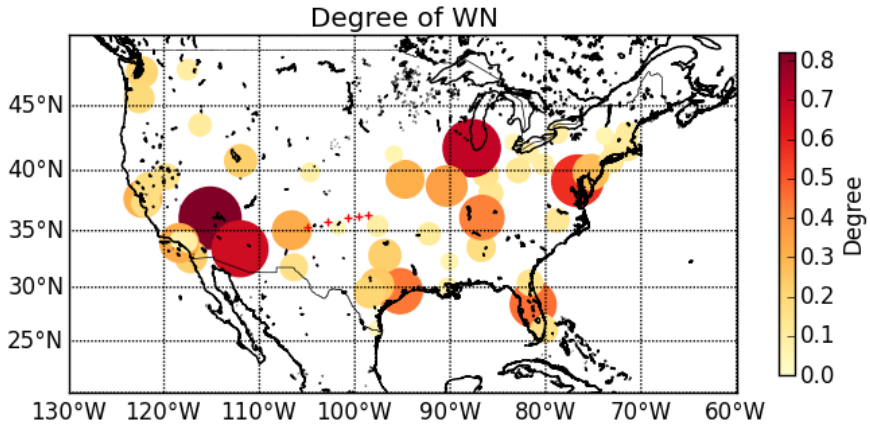

Figure 18: *Degree centrality for Southwest airlines in 2013, including some COGs over time.*

illustrates the moving COG's displayed as crosses. The red cross on the far left is the COG in the first quarter of 2000 and the most right of 2013. To construct these variables including passengers and geographical location of airports, the difference in distance to the COG is calculated over 2 successive quarters. The first quarter of the data set, however, has no difference to compute, lacking a model variable. For example, as an airport's distance from the COG is increased, the resulting variable will have a negative effect on the betweenness centrality of that airport for an airline. This is clearly visible for Southwest as

flights are being moved from the east to west coast.

6 Conclusions and Further Work

Throughout this work, we have first seen with an econometric study that the behaviour of carriers evolves in time as do their pricing policies. This evolution could be the consequence of some one-off events like a change of CEO, or a more general tendency, to meet the demand of an airline market in a permanent change, with the expansion of low-cost carriers or the merging of legacy ones, for example. Moreover, in this environment more and more competitive, carriers give a huge importance to their network and adopt different expansion strategies.

In the forecasting of these networks, some steps we could take to go further are first to use older values than the first lagged values of the explanatory variables. Indeed, it certainly takes more time than a quarter for an airline to open new routes and to adapt to some changes of the market. For example, if the number of passengers travelling from an airport significantly increases in a quarter, the airline would certainly rather wait a few quarters and see if this fact is not a singular event before opening new routes from this airport. Then, we should integer other airlines network in order to see if their is an interdependence between them, as in venture capital networks [6].

Another point is that we chose to gather low-cost and legacy carriers in our study, but we could have done this separately insofar as the strategies and the services offered by these two types of airlines are mostly different.

The final point would be to use the network forecast properties in order to improve the model of the mean real fare of a carrier on a route by forecasting method. Although the centrality model produces reasonable results, it can be improved using variables for different airline strategies and networks. Many networks have been seen to combine multiple networks, which at this stage are modelled simultaneously. For example, some legacy airlines make use of two or three star shaped networks instead of one. Southwest as well, when examining routes separately, contains many small structures in which changes occur locally, rather than modelling it at continental level. The last section mentioned a new variable regarding the centre

of gravity calculated with routes and passengers. As specified before, the capacity and exclusivity of an airport play a significant role in centralities and air fares. Not all variables can be extracted from the data used, but might be helpful for the estimation.

References

[1] A. Dossin and S. Lawford, "Weighted networks with application to U.S. domestic airlines." Presentation available at: http://leea.recherche.enac.fr/Steve%20Lawford/presentations/DEVI%20ENAC%2022%20Sep%202016.pdf, 2016.

[2] G. Cook and J. Goodwin, "Airline networks: A comparison of hub-and-spoke and point-to-point systems airline networks," *Journal of Aviation/Aerospace*, pp. 33–42, 2008.

[3] J. Morancé and T. Noortman, "An empirical investigation of network structure: Evidence from U.S. domestic airlines." 2014.

[4] S. Borenstein, "Hubs and high fares: Dominance and market power in the u.s airline industry," *The RAND Journal of Economics*, pp. 344–365, 1989.

[5] C. Oster and J. Strong, "Predatory practices in the U.S. airline industry." Paper available at: http://ntl.bts.gov/lib/17000/17600/17602/PB2001102478.pdf, 2001.

[6] V. Hochberg, A. Ljungqvist, and Y. Lu, "Whom you know matters: Venture capital networks and investment performance," *Journal of Finance*, pp. 293–295, 2007.

A Appendices

Dependent Variable: MEANREALFARE
Method: Least Squares
Date: 10/09/16 Time: 14:00
Sample: 1 1623
Included observations: 1582
Weighting series: PAX^0.5
Weight type: Variance (average scaling)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	168.7340	13.21544	12.76795	0.0000
MINBETWEENNESS	227.7939	75.94500	2.999459	0.0027
MAXBETWEENNESS	28.08219	6.729013	4.173300	0.0000
DISTANCE/100	20.32093	1.177143	17.26292	0.0000
(DISTANCE/100)^2	-0.361718	0.043707	-8.276047	0.0000
ABSTEMPDIFF	-0.781050	0.247017	-3.161933	0.0016
MEANGDPPERCAPITA	0.231465	0.185920	1.244967	0.2133
T100SEATS/100000	8.504633	3.135541	2.712334	0.0068
MONOPOLY	32.06440	3.834310	8.362496	0.0000
COMPETITIVE	-18.95770	5.030078	-3.768868	0.0002
AA	26.96916	6.490961	4.154879	0.0000
AS	-39.88045	11.80480	-3.378325	0.0007
B6	-13.04644	7.504531	-1.738476	0.0823
DL	123.4473	5.461630	22.60264	0.0000
F9	-114.3161	9.449741	-12.09728	0.0000
FL	-14.49986	9.726398	-1.490774	0.1362
NK	-197.5546	7.784858	-25.37678	0.0000
SY	-52.97036	16.45975	-3.218175	0.0013
UA	84.87648	6.445545	13.16824	0.0000
US	88.04881	7.883537	11.16869	0.0000
VX	28.05481	16.23084	1.728487	0.0841
AA*WNPRESENCE	3.380581	25.95555	0.130245	0.8964
AS*WNPRESENCE	-2.298821	24.70646	-0.093045	0.9259
B6*WNPRESENCE	-4.003545	17.78336	-0.225129	0.8219
DL*WNPRESENCE	-84.10838	12.15871	-6.917542	0.0000
F9*WNPRESENCE	38.21325	15.72099	2.430715	0.0152
FL*WNPRESENCE	3.623277	11.70072	0.309663	0.7569
NK*WNPRESENCE	-0.203963	16.51880	-0.012347	0.9902
SY*WNPRESENCE	-41.76596	39.93367	-1.045883	0.2958
UA*WNPRESENCE	-54.89843	11.31535	-4.851676	0.0000
US*WNPRESENCE	-93.28281	12.66313	-7.366488	0.0000
VX*WNPRESENCE	-82.92526	39.06631	-2.122679	0.0339

Weighted Statistics

R-squared	0.758961	Mean dependent var	362.7355
Adjusted R-squared	0.754140	S.D. dependent var	147.6240
S.E. of regression	60.33925	Akaike info criterion	11.05786
Sum squared resid	5643280.	Schwarz criterion	11.16641
Log likelihood	-8714.770	Hannan-Quinn criter.	11.09819
F-statistic	157.4354	Durbin-Watson stat	2.083493
Prob(F-statistic)	0.000000	Weighted mean dep.	372.2324

Unweighted Statistics

R-squared	0.767760	Mean dependent var	370.4834
Adjusted R-squared	0.763115	S.D. dependent var	118.6646
S.E. of regression	57.75509	Sum squared resid	5170258.
Durbin-Watson stat	2.115901		

Figure A.1: *Baseline static regression: 2013Q4.*

Dependent Variable: BETWEENNESS
Method: Least Squares
Date: 01/03/17 Time: 14:41
Sample: 1999Q1 2011Q4 IF CARRIER="AA" AND AIRPORT="MIA"
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002166	0.005068	0.427396	0.6711
BETWEENNESS(-1)	0.705399	0.109864	6.420658	0.0000
MONOPOLYROUTES(-1)	2.38E-05	0.000603	0.039400	0.9687
COMPETITIVEROUTES(-1)	3.90E-05	0.001603	0.024326	0.9807
PAX(-1)	2.88E-07	2.48E-07	1.165371	0.2499

R-squared	0.686046	Mean dependent var	0.035853
Adjusted R-squared	0.658745	S.D. dependent var	0.010215
S.E. of regression	0.005967	Akaike info criterion	-7.312089
Sum squared resid	0.001638	Schwarz criterion	-7.122694
Log likelihood	191.4583	Hannan-Quinn criter.	-7.239715
F-statistic	25.12954	Durbin-Watson stat	1.982892
Prob(F-statistic)	0.000000		

Figure A.2: *First forecast model.*

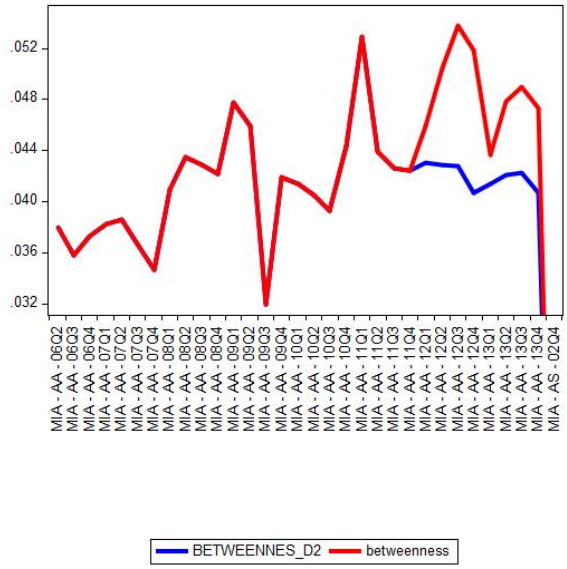

Figure A.3: *Forecast results.*

Dependent Variable: BETWEENNESS
Method: Least Squares
Date: 01/06/17 Time: 13:46
Sample: 1999Q1 2011Q4 IF AIRPORT="MIA" AND CARRIER="AA"
Included observations: 51

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.060358	0.031683	1.905056	0.0635
BETWEENNESS(-1)	0.443870	0.125194	3.545448	0.0010
MONOPOLYROUTES(-1)	-0.001048	0.000770	-1.360206	0.1809
COMPETITIVEROUTES(-1)	0.000694	0.001545	0.449285	0.6555
PAX(-1)	-7.11E-08	2.67E-07	-0.265983	0.7915
CAPACITY(-1)	1.44E-08	6.43E-09	2.238328	0.0304
LOADFACTOR(-1)	-0.029148	0.024977	-1.166982	0.2496
NODES(-1)	-0.000599	0.000242	-2.474738	0.0174

R-squared	0.755473	Mean dependent var	0.035853
Adjusted R-squared	0.715666	S.D. dependent var	0.010215
S.E. of regression	0.005447	Akaike info criterion	-7.444363
Sum squared resid	0.001276	Schwarz criterion	-7.141331
Log likelihood	197.8313	Hannan-Quinn criter.	-7.328566
F-statistic	18.97852	Durbin-Watson stat	2.168568
Prob(F-statistic)	0.000000		

Figure A.4: *Second forecast model.*

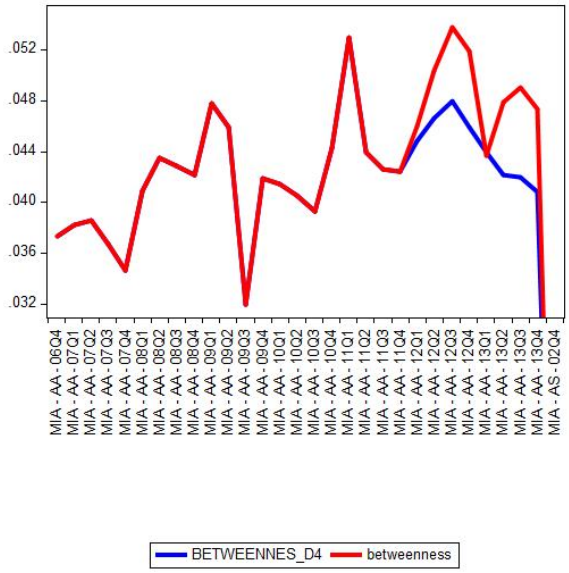

Figure A.5: *Forecast results.*

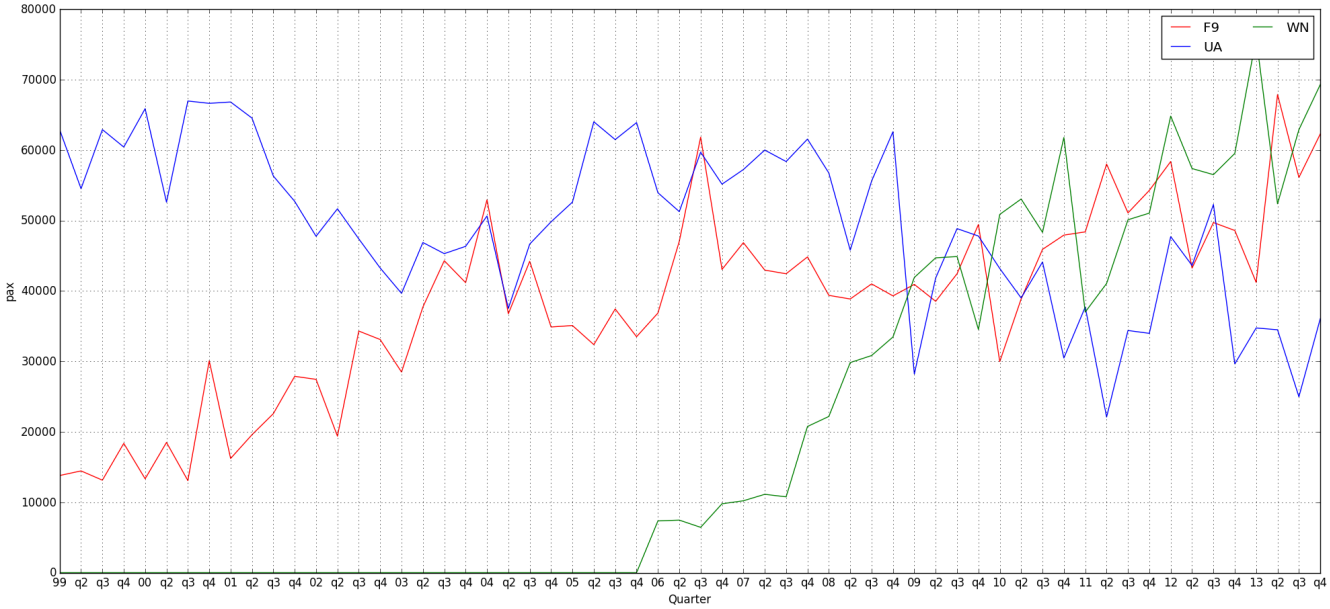

Figure A.6: *Passenger numbers for every airline using Denver as a hub.*