Replication report
Tracing the Woes: An Empirical Analysis of the Airline Industry

Wiktor Zielinski

December 3, 2023

1 Most important finding of the paper

The authors try to search for an explanation for the complicated facts: in the early 2000s, the stats showed that the airline industry was developing - domestic revenue passenger miles were gradually increasing same as the load factor. However, only during the studied period of 1999-2006, there were 8 bankruptcies, including the biggest companies such as US Airways, United Airlines, Delta and Northwest.

The authors argue that the demand shocks: 2001 dot-com bubble, 9/11, development of the search engines; as well as supply shocks: entry of Low-Cost Carriers (LCC), progress in aviation technology affected the structure of demand.

The main findings of the paper on this structural change are the following:

- Compared to 1999, the price elasticity of air travel demand increased by 8 percent.
- Connection flights became much less preferable - the connection semi-elasticity was 17 percent higher.
- Due to increasing fuel price and progress in aviation technology, changes in marginal cost favored non-stop flights.

The decline in variable profits among legacy carriers, accounting for over 80%, is attributed to the combination of factors mentioned above, including the rise of low-cost carriers. More than half of this reduction can be traced back to shifts in demand.

2 Replication

As the authors use Matlab specific functions, I decided not to change the environment to replicate the study. First, we need to see that there is a C code being used in the analysis. Therefore, it is needed to compile the C function before using the program. We do it by typing a command "mex M40_MkSum.c". Second, we have to turn off the setting of 'DerivativeCheck' from 'on' to 'off' in the optimset function. Otherwise, Matlab compiler returns an error and stops the analysis. We assume this error is caused by numerical precision imposed by Matlab, as changing the tolerance level did not fix the issue. However, we trust that the provided gradient is correct and we manage to obtain exactly the same results as the authors did.

Now, I will briefly describe the replication process. First, the authors proceed with the IV estimation - they instrument prices from demand and supply side, just like in the paper. They later input an inverted matrix of the self-product of instruments into an object VM, to include them in the GMM estimation. Then, they proceed to estimation, setting up starting values of parameters, lower and upper bounds for the parameters (mainly to speed up the optimization). Our optimization is unconstrained and subject only to the lower and upper bounds provided. Xmat is vector of observable characteristics used in the analysis. Importantly, dM.modspec sets a scenario to be estimated - 68 provides us with basic scenario for estimates. All the possible scenarios are visible in M130_xbMM.m.

After estimating the first stage, we proceed to the second stage estimation, improving our previously estimated parameters and including the supply side. All the analysis is taking place inside the gradient function M130_gmmGredMM which refers to all the other functions provided - a short description of the functions can be found in ReadMe file.

A technical note worth mentioning is that all the analysis differs for 1999 and 2006 mainly because we have different airlines in these periods. That's why the functions are sometimes year specific (starting with P or M), sometimes the year issue is treated inside the functions.

We manage to replicate the exact results of the analysis provided in the paper, to be found in Table 1.

3 Assumption modification

Given that the authors supply us with the data already after sample selection process, there is not much space to modify assumptions related to this part. However, I find the definition of a market a bit complicated. It is a directional pair of an origin and a destination airport. However, I see no reason to put it in that way. First, it is impossible for a carrier to fly only one-way. Second, to attract demand and use fully consumer preferences the carrier can 'smooth' the prices using the round-trip. I decide to collapse the provided dataset. All the data manipulation is conducted in 'collapse.R' file.

To be able to identify and join the round-trip markets, I have to have a unique identifier of my new market. The distance provided by the authors is not the distance between the cities, but a trip distance. Therefore, I construct a new variable that takes the minimum distance in a given market and assigns it to every row in a market. On that basis, I create a new market index.

Then, I change the definition of a product from one-way trip to round-trip. To do that, I use the original distance. That ensures me to precisely define the product - a flight between 2 cities that has connection is matched with the same connected trip but in other direction. In that way, I can still retrieve connection semi-elasticity in the analysis. In the collapsing process, I take mean of the all variables except dummies to avoid colinearity - I take min and max that reflect parameters of the whole round trip. For example, $min = max = 1$ for the hub variable means that both, the origin and the destination is a hub for the carrier. I proceed in a similar way for numerical destination and origin variables - to reflect the new market I take the mean, but for the other field I have to take some other value (in my case min). That's a caveat of this specification. However, it should not have huge implications for the results as it is still a route characteristics. For dummies that are not origin-destination specific, I take a max value.

For 1999, number of products reduces from 214,809 to 125,731 while number of markets from 3,998 to 1,442. For 2006, from 226,532 to 130,706 and from 4,300 to 1,538.

The changes in the parameter estimates can be seen in Table 1. Our main point of the analysis, price elasticity enforces the previous finding - the change between 1999 and 2006 for tourists now is 72% comparing to 27% of the original. For 1999, we observe no change in price elasticity for business clients, but again, a large difference in 2006. This might reflect that businesses started to regulate travel spending and used search engines as well.

Second main point is connection semi-elasticity. We see almost no change in the base result of 1999, with a more significant change in 2006. Again, the results confirm the previous intuition - connected flights are unfavorable to direct flights, in the modified case mainly for the tourists.

On the supply side, we can see much higher coefficient for connection short and long in 2006 with large standard errors. This is probably caused by a misspecification when collapsing the data, taking max value does not reflect the whole round trip characteristics. Perhaps we could think of other instruments better reflecting the supply side in that case. Beside that, the trends differ slightly for distance. However, I would not put too much attention on interpreting the supply side. There seems to be an issue with our collapsing move that needs a different specification.

Finally, we observe a different trend in lambda - our new data shows that clients became more attached to the career which does not seem in line with the narration and new LCC data. However, JetBlue and SouthWest are often counted as LCC and our estimates may have captured this. Share of tourists follows the same, a bit sharper change.

To sum up, the modification implemented still reflects the trends highlighted in the paper. However, it increases our estimates for the main findings even more - in the new scenario we can see higher price elasticity change and preference for direct flights. Nevertheless, the supply side seems to be misspecified - a problem which would need to be addressed in a broader analysis.

	1999 modified	1999 original	2006 modified	2006 original
fare1	−0.70	−0.78	−1.42	−1.05
	(0.00)	(0.02)	(0.01)	(0.03)
connection1	−0.53	−0.53	−0.65	−0.59
	(0.00)	(0.02)	(0.01)	(0.03)
const1	−6.02	−5.79	−7.06	−5.68
	(0.06)	(0.19)	(0.42)	(0.19)
fare2	−0.07	−0.07	−0.16	−0.10
	(0.00)	(0.00)	(0.00)	(0.00)
connection2	−0.34	−0.31	−0.39	−0.50
	(0.00)	(0.02)	(0.00)	(0.02)
const2	−8.69	−8.55	−7.63	−8.61
	(0.11)	(0.40)	(0.36)	(0.30)
ndest	0.36	0.38	0.47	0.27
	(0.02)	(0.03)	(0.02)	(0.02)
ndept	0.05	0.04	0.06	0.11
	(0.00)	(0.00)	(0.00)	(0.00)
dist	0.26	0.30	0.18	0.53
	(0.03)	(0.04)	(0.03)	(0.04)
dist2	−0.04	−0.05	−0.04	−0.08
	(0.01)	(0.01)	(0.00)	(0.01)
tour	0.34	0.30	0.32	0.36
	(0.03)	(0.03)	(0.03)	(0.03)
slot_control	−0.21	−0.19	−0.13	−0.18
	(0.01)	(0.01)	(0.01)	(0.01)
const_short	0.86	1.07	0.93	1.17
	(0.63)	(0.06)	(0.96)	(0.06)
dist_short	0.25	0.26	0.61	0.19
	(0.60)	(0.01)	(0.80)	(0.01)
connection_short	−0.07	−0.06	−0.76	0.06
	(0.20)	(0.03)	(0.27)	(0.04)
const_long	1.39	1.62	1.27	1.59
	(0.52)	(0.08)	(0.75)	(0.07)
dist_long	0.08	0.09	0.15	0.04
	(0.11)	(0.01)	(0.15)	(0.01)
connection_long	−0.09	−0.09	−0.85	0.05
	(0.08)	(0.03)	(0.12)	(0.04)
hubMC	0.00	−0.02	−0.09	−0.05
	(0.27)	(0.01)	(0.50)	(0.01)
slotMC	0.06	0.08	−0.77	0.03
	(0.17)	(0.01)	(0.27)	(0.01)
lambda	0.79	0.77	0.80	0.72
	(0.00)	(0.01)	(0.00)	(0.01)
gamma	0.63	0.70	0.45	0.63
	(0.04)	(0.12)	(0.19)	(0.11)

Table 1: Base case parameter estimates - original and modified