

PUI2015 Extra Credit Project

Oil Price and Air Fare

<Jiheng Huang, <https://github.com/JihengHuang>, jh5138>

This project focuses on investigating that whether air passengers benefit from the recent decline in oil price.

The conclusions are: 3 City pairs: Los Angeles and San Francisco, Atlanta and Chicago, Atlanta and Orlando. Their air fares are **highly correlated** to oil price. 1 City pair: Chicago and Los Angeles: air fare **does not have** obvious correlation with oil price (high p value). Considering lag effect **does not change** much on spearman correlation results.

Data:

Air Fare Data: Domestic Airline Consumer Airfare Report from U.S. Department of Transportation, Table 6 – Detailed Fare Information for All City-Pair Markets. Data is quarterly collected from year 1996 1st quarter to 2015 1st quarter. (Link:

<https://www.transportation.gov/policy/aviation-policy/domestic-airline-consumer-airfare-report>)

Oil Price Data: West Texas Intermediate (WTI) Oil Price. Data is **averaged** quarterly from 1996 1st quarter to 2015 1st quarter (same data period as air fare data). Before using python code to read the excel(.xls) data file, a brief introduction of the data in the first 10 rows was deleted so the dataframe would only contain timestamp and price data. (Link:

<https://research.stlouisfed.org/fred2/series/DCOILWTICO/downloaddata>)

Top 10 City Pairs:

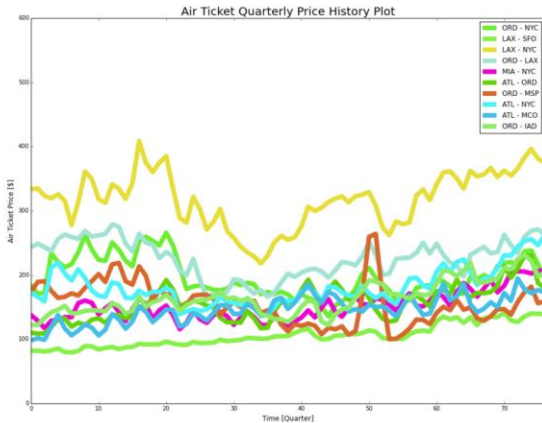
Top 10 popular city pairs vary in different years and different quarters. So Wikipedia definition is used here:

Busiest 10 air routes by city pairs within the United States on Wikipedia (September 2014 - August 2015 data):

(Link: https://en.wikipedia.org/wiki/World%27s_busiest_passenger_air_routes)

Median 10 pairs on popularity and the least 10 popular routes were not considered also because city pairs change from time to time. This project focused on selected 10 popular routes.

missing caption and label WAAAAAAAY too small!!!



you should use the same colors when you plot the same thing:
the plot to the right should use the same color per city pair as the plot on
the left, and i would have chosen thin lines over semi-transparent lines

Methodology:

First, air fare of the 10 busiest air routes in the United States over 20 years was plotted (data published quarterly, 1996 1st quarter to 2015 1st quarter. $(2015 - 1996) * 4 + 1 = 77$ data for each route)

since transparency is hard to judge, you cannot use it as a discriminator
in the caption. you can use thin vs thick lines though.

Second, Air fare (transparency = 0.5) and oil price (red line) at the same time were plotted to give a brief idea of correlation visually.

Third, Spearman correlation test was conducted on the time-series data of air fare and oil price.

Fourth, Lag effect was considered. 2 situations: (1) Air fare change is lagging 1 quarter behind oil price change (2) Air fare change is lagging 2 quarters behind oil price change were considered and reran the spearman correlation test

since you do not provide a legend to go from airport code to city name at least
for the 3 main pairs you should have spelled it out (for example i do not know
what airport code orlando corresponds to)

Conclusions:

3 City pairs: Los Angeles and San Francisco, Atlanta and Chicago, Atlanta and Orlando. Their air fares are **highly correlated** to oil price.

1 City pair: Chicago and Los Angeles: air fare **does not have** obvious correlation with oil price (high p value).

Considering lag effect **does not change** much on spearman correlation results.

Spearman Correlation test result about Air Fare and Oil Price (No lag):

City1	City2	Correlation Coefficient	p-value
Chicago	New York City	-0.28	0.01
Los Angeles	San Francisco	0.91	0

Los Angeles	New York City	0.23	0.04
Chicago	Los Angeles	-0.12	0.29
Miami	New York City	0.57	0
Atlanta	Chicago	0.70	0
Chicago	Minneapolis	-0.54	0
Atlanta	New York City	0.40	0
Atlanta	Orlando	0.75	0
Chicago	Washington DC	0.65	0

Spearman Correlation test result about Air Fare and Oil Price (1 quarter lag):

City1	City2	Correlation Coefficient	p-value
Chicago	New York City	-0.23	0.05
Los Angeles	San Francisco	0.92	0
Los Angeles	New York City	0.28	0.01
Chicago	Los Angeles	-0.06	0.60
Miami	New York City	0.61	0
Atlanta	Chicago	0.69	0
Chicago	Minneapolis	-0.45	0
Atlanta	New York City	0.43	0
Atlanta	Orlando	0.74	0
Chicago	Washington DC	0.68	0

Spearman Correlation test result about Air Fare and Oil Price (2 quarters lag):

City1	City2	Correlation Coefficient	p-value
Chicago	New York City	-0.24	0.04
Los Angeles	San Francisco	0.91	0
Los Angeles	New York City	0.27	0.02
Chicago	Los Angeles	-0.04	0.74
Miami	New York City	0.63	0
Atlanta	Chicago	0.67	0
Chicago	Minneapolis	-0.41	0
Atlanta	New York City	0.44	0
Atlanta	Orlando	0.74	0
Chicago	Washington DC	0.68	0

Future work: More statistical tests: for instance, lag regression can be conducted. More routes can be selected. Different pricing pattern on different airlines can also be studied.

Links:

Code: https://github.com/fedhere/PUI2015_EC/blob/master/jh5138_EC/AirFare.ipynb