Technical Note DiD Case Study

Sakib Anwar

April 3, 2025

1 Introduction

This documents explains the python notebooks used for the DiD Case Study. Part 1 of the document is about Data Preparation and Part 2 of the document is about Analysis.

2 Part 1 : Data Preparation

This technical note explains the full sequence of data processing steps in the Jupyter notebook `airlines-01-dataprep.ipynb`. The objective of the notebook is to clean, transform, and prepare a panel dataset of airline routes before and after the merger between **American Airlines (AA)** and **US Airways (US)** for further analysis—potentially a *difference-in-differences (DiD)* or panel regression study.

1. Setup and Imports

- Imported standard libraries: `numpy`, `pandas`, `os`, `sys`, `warnings`.
- Suppressed warnings for cleaner output.

2. Load Dataset

The dataset is loaded from an OSF-hosted Stata file:

```
data = pd.read_stata("https://osf.io/zw2h9/download")
```

3. Filter for Analysis Years

Only the years 2011 (pre-merger) and 2016 (post-merger) are retained for the analysis.

4. Construct Total Passengers by Carrier

Converted share variables to absolute passenger counts by multiplying shares with total passengers:

- `ptotalAA = shareAA * passengers`
- `ptotalUS = shareUS * passengers`
- `ptotallargest = sharelargest * passengers`

5. Aggregate to Market-Year Level

Grouped by origin, final destination, return status, and year. Aggregated passenger counts, fares, and derived totals.

6. Time Indicators

Created binary variables `before` and `after` to indicate pre- and post-merger observations.

7. Compute Market-Level Metrics

- `avgprice`: average fare
- `shareAA`, `shareUS`, `sharelargest`
- Dummies for presence of AA, US, both, or either.

8. Market Identifier

Constructed a unique string ID for each market using origin, destination, and return indicators.

9. Compute Passenger Counts Before and After

Identified and averaged passengers before and after merger by market.

10. Balanced Panel Identification

Markets appearing in both 2011 and 2016 were marked as `balanced = True`.

11. Balanced Market Summary

Summarized passengers and market counts for balanced vs unbalanced observations.

12. Treatment and Market Type Flags

Defined three flags based on pre-merger characteristics:

- `treated`: market had both AA and US in 2011.
- `untreated`: market had neither.
- `smallmkt`: fewer than 5,000 passengers.

13. Log and First Difference of Prices

Computed:

- Log of average price (`lnavgp`) while handling `inf`.
- First difference in log prices (`d_lnavgp`) within each market.

14. Descriptive Statistics

Produced summaries for:

- Yearly passengers
- JFK–LAX market
- Small vs large markets
- Balanced vs unbalanced panels
- Treated vs untreated markets
- Prices with zero values

15. Save Cleaned Dataset

Saved the processed dataset to disk using:

```
data_agg.to_pickle("airline-workfile.pkl")
```

16. Export Descriptive Table to LaTeX

Generated a LaTeX table from transposed summary statistics:

```
latex_output_all = descriptive_stats_all.T.to_latex()
```

Conclusion

This notebook prepares a robust panel dataset suitable for evaluating the competitive effects of airline mergers. Key features include:

- Balanced panel filtering
- Time-based and market-specific indicators
- Pre-treatment classification of markets
- Descriptive and diagnostic checks
- Log transformation and differencing of prices
- Export of a clean dataset and LaTeX summary table

3 Part 2 : Data Analysis

This technical note explains the data analysis workflow used to assess the impact of the American Airlines–US Airways merger using a *difference-in-differences (DiD)* approach and *weighted least squares (WLS)* regression. The notebook uses pre-processed panel data on airfares across markets and time periods.

1. Setup and Imports

- Key libraries: `pandas`, `numpy`, `statsmodels`, `plotnine`, etc.
- `plotnine` is used for ggplot-style plots.
- Warnings are suppressed and helper functions are defined.

2. Load and Merge Datasets

- Re-loads raw dataset from OSF and merges with market-level data.
- Adds flags: `treated`, `smallmkt`.

3. Define Weighted Averages

- Custom lambda function calculates weighted averages using passenger numbers.

4. Average Price Trends Over Time

- Aggregates `avgprice` and computes `lnavgprice`.
- Plots log average fare trends by treatment status for:
  - All markets
  - Small markets
  - Large markets

5. Visualizing Trends

- Uses `plotnine` to generate line plots.
- Visual inspection of parallel trends assumption.

6. Run First DiD Regression

- WLS model: `d_lnavgp` $\sim$ `treated`
- Weights: pre-merger passengers
- Separate models for all, small, and large markets.

7. Stargazer Output

- Uses `Stargazer` to format regression outputs.
- Adds readable column labels and renames coefficients.

8. Conditioning on Observables

- Adds controls: `lnpass_bef`, `return_`, `stops`, `sharelarge_bef`.
- New model: `d_lnavgp` $\sim$ `treated + controls`

9. Treatment Intensity

- Replaces binary treatment with `share_bef`.
- Plots histogram of treatment intensity weighted by passengers.
- Regression: `d_lnavgp` $\sim$ `share_bef + controls`

10. Conditional DiD Specification

- Fully interacted DiD model:
- `lnavgp` $\sim$ `(treatment + controls) * after`

11. Market Summary Tables

- Reports passengers by `balanced`, `before`, and `treatment_isna`.

12. Coefficient Plot Functions

- Functions visualize coefficient estimates and 95% CI across models.
- Facilitates comparison between all/small/large markets.

13. Output: Coefficient Comparison

- Coefficients plotted for:
  - Binary treatment (`treated`)
  - Continuous treatment (`share_bef`)
  - Fully interacted model

Summary

- Balanced panel analysis of airline markets.
- Multiple DiD specifications with robustness checks.
- Weighted regressions and visualization support inference.
- Ready-to-export LaTeX tables and plots for publication.