UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

THREE ESSAYS ON THE U.S. AIRLINE INDUSTRY

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

ZEXUAN LIU
Norman, Oklahoma
2016

THREE ESSAYS ON THE U.S. AIRLINE INDUSTRY

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF ECONOMICS

BY

Dr. Qihong Liu, Chair

Dr. Scott Linn

Dr. Pallab Ghosh

Dr. Myongjin Kim

Dr. Georgia Kosmopoulou

Dr. Cynthia Rogers

I dedicate this dissertation to my parents, Xingjian Liu and Huiying Zhu, for their kindness and devotion; my mentor, Yanyuan Fu, for giving me the inspiration to aim high and the guidance to achieve my dreams.

Acknowledgements

I would like to acknowledge the guidance and support of my committee members: Qihong Liu, Pallab Ghosh, Cynthia Rogers, Myongjin Kim, Georgia Kosmopoulou, and Scott Linn. Furthermore, my fellow cohorts, Sun Hyung Lee, Jimmy Schlaffer, Weici Yuan, and Lee Stitzel, have made substantial contribution to this endeavor. I was fortunate to be surrounded by these people and I cannot thank them enough for all of their encouragement. I also appreciate financial and travel support from the OU Economics department.

Table of Contents

Acknowledgements . . . . **iv**

List of Tables . . . . **viii**

List of Figures . . . . **ix**

Abstract . . . . **x**

1 An Application of Genetic Matching on the US Airlines Merger 1

1.1 Introduction 1

1.1.1 Literature review 3

1.2 Econometric Model 6

1.2.1 Model Setup 6

1.2.2 Propensity Score Matching 8

1.2.3 Mahalanobis Distance 10

1.2.4 Genetic Matching 11

1.3 Application to US Airways – America West merger 14

1.3.1 Data and preliminary estimation results 14

1.3.2 Market characteristics used for matching 19

1.3.3 "Control" group: Matched vs. Comparison group 20

1.3.4 Another look at the data 28

1.4 Conclusion 30

2 Is the Tarmac Delay Rule a Panacea for Passenger Right Protection? **32**

2.1 Introduction 32

2.2 Background 35

2.3 Data 38
2.3.1 The importance of addressing network effects 38
2.3.2 Measuring the Delay Increase 41
2.4 Estimation strategy 42
2.4.1 The impact of policy on cancellation 42
2.4.2 The validity of diff-in-diff method 43
2.4.3 Rearrangement efficiency 47
2.5 Results 48
2.5.1 Baseline analysis 48
2.5.2 Differential impacts on legacy vs. regional carriers 50
2.5.3 Rearrangement efficiency 52
2.6 Falsification Test 54
2.7 Conclusion, Policy Implication and Discussion 57
3 Product Differentiation and Quality Uncertainty in the U.S. Airline Industry 59
3.1 Introduction 59
3.2 Theories and empirical studies of horizontal differentiation 61
3.2.1 Theories of horizontal differentiation and quality uncertainty 61
3.2.2 Empirical studies of horizontal differentiation in airline industry 63
3.3 Data and measures of temporal differentiation 64
3.3.1 Data 64
3.3.2 Measures of temporal differentiation 66
3.3.3 Own flight v.s. one-sided flight v.s. two-sided flight 68
3.3.4 Summary statistics 69

3.4 Estimation strategies 73

3.5 Results 75

3.5.1 Baseline estimation result 75

3.5.2 Time to noon across different market structure 78

3.5.3 Impacts by different flight types 79

3.5.4 Robustness tests 82

3.6 Conclusion 85

References **90**

Appendix **91**

3.7 Instruction on how the data is cleaned and organized 91

List of Tables

1.1 Summary Statistics for Airline Merger 17

1.2 Summary Statistics All Pre-merger Periods for Airline Merger . . 22

1.3 Common Trend Comparison: Simple DiD vs. DiD Matching(9 lags) vs. DiD Matching(4 lags) 23

1.4 Falsification Test for Airline Merger 26

1.5 Results Comparison: OLS vs. Simple DiD vs. DiD Matching(9 lags) vs. DiD Matching(4 lags) 27

2.1 DOT bills list 37

2.2 Summary Statistics for Tarmac Delay Rule Study 39

2.3 Difference-in-difference Common Trend Test for Tarmac Delay Rule Study 46

2.4 Estimation Results of Tarmac Delay Rule Study 51

2.5 Cancellation Ratio and Cancellation Multiplier 54

2.6 Falsification Test for Tarmac Delay Rule Study 56

3.1 Summary statistics 71

3.2 Baseline results 76

3.3 Time to noon results 80

3.4 Gini coefficient results 81

3.5 Robustness check of time to noon 83

3.6 Robustness check of Gini coefficient 84

List of Figures

2.1 Common Trend Comparison for Hub and Non-hub Airports . . . 44

2.2 Average Number of Flights Comparison over Time 52

2.3 Flights with Re-shuffle Behavior 55

3.1 Histogram of Gini Coefficient by Flight Type 72

Abstract

My research uses cutting-edge econometric techniques to isolate and identify the impact of various changes in the U.S. airline industry, whether coming from an endogenous change in the market structure such as merger, or coming from a (potentially) exogenous policy change such as the tarmac delay rule. My research is primarily empirical and I have utilized large microeconomic data from the airline industry.

In my first chapter, I use genetic matching and difference-in-difference to estimate the impact of airline merger on vertical and horizontal product differentiation. Different from many industries, airline services can be viewed as networks. Correspondingly, merger can lead to significant network effects, in addition to the commonly claimed synergy (economies of scale). There have been several existing studies trying to link market structure with on-time performance and my paper adds to this literature. The main contributions of our paper are three-folds. First, I focus on one particular merger (between U.S. Airways and America West) and carefully select sample periods to avoid contamination of confounding factors. Second, I use panel matching together with endogenous weights to construct a properly matched group: those that did not experience merger but best resemble the treated group (from merging carriers). Comparing the treated and matched groups allows us to do a proper difference-in-difference estimation to evaluate the impact of merger. Third, and probably most important, while I analyze the impact of merger on arrival delay, I take it one step further by treating arrival delay as one measure of quality (vertical product differentiation). I then introduce several other measures for vertical differentiation as well as horizontal differentiation. To my knowledge, this is the first study of the U.S. airline

industry taking this perspective.

To estimate the causal impact of merger, I use panel matching with a difference in difference approach. Maintaining the time series property is important in terms of constructing a properly matched group, especially for the merger case. To address this, I use the genetic algorithm to find the optimal weight for each covariate per time period. Based on the optimal weights, a generalized Mahalanobis distance matrix is created. I construct the matched group by identifying the panels with a minimum distance to the treated panels. Conditional on the balanced covariates, I are able to address the selection issue and then use a diff-in-diff estimation to gauge the impact of merger. Our results show that merger reduces arrival delay (by 0.7 minutes), number of flights per route (by 10.7), number of routes (by 61) and destinations (by 2.8). On the other hand, merger has mostly insignificant impacts on our measures of horizontal differentiation.

My second chapter's topic is motivated by both personal experience and the extensive media coverage about the lengthy tarmac delay. As a response to multiple lengthy tarmac delay cases, U.S. Dept of Transportation (DOT) released a new tarmac delay rule in December 2009. The purpose of the rule is to eliminate the lengthy tarmac delay and protect consumers' rights. Carriers face a severe financial penalty if they fail to provide opportunities for passengers to deplane after a three-hour tarmac delay for domestic flights. This paper aims to examine the impact of the tarmac delay rule, including carriers' responses to the rule.

My third chapter studies the impact of quality uncertainty on flight temporal differentiation. There is an extensive literature studying optimal horizontal product differentiation. It is well documented that when a firm moves its product closer to its rival's product space, there are two opposite effects on the firm's profitability: market share effect which improves profit and competition effect (or price effect) which reduces profit. More recently, Bester (1998) links the impact of quality uncertainty with the pattern of equilibrium horizontal differentiation. Their main finding is that quality uncertainty reduces the intensity of the price effect, resulting in lower horizontal differentiation in the equilibrium.

Using two measures of horizontal differentiation: Time to noon and Gini coefficient, I directly test this hypothesis. After using a novel instrumental variable to correct for endogeneity, I find that higher quality uncertainty is linked to less horizontal differentiation, supporting Bester's theoretical prediction. With higher quality uncertainty, travelers do not view firms' flights as close substitutes even though their departure times may be close. This reduces competition intensity (the price effect). As a result, the market share effect will dominate the price effect and results in an "aggregation" in terms of horizontal differentiation.

Chapter 1

An Application of Genetic Matching on the US Airlines Merger

1.1 Introduction

Airline industry is a significant part of the U.S. economy. In 2012, it accounted for "5.4% of our gross domestic product (GDP), contributed $1.5 trillion in total economic activity, and supported 11.8 million jobs" (FAA 2014 economic impact report). The U.S. airline industry has experienced significant consolidation in the last decade or so. It is important in both theory and practice to understand the impact of such consolidation. Much of the focus in existing studies has been on pricing, in particular, the relationship between competition and price dispersion, and the results have been mixed.[1] Also commonly studied is airlines' on-time performance. Rupp, Owens and Plumly (2001) and Mazzeo (2003) both look at the link between on-time performance and competition but reach different conclusions. Prince and Simon (forthcoming) consider 5 recent mergers in the airline industry and investigate how these mergers affect on-time performance. They find that merger has a negative (worse on-time performance) short-term impact, but no or positive long-term impact.

In this paper, we study the impact of one particular merger – that between U.S. Airways and America West which took place in 2005. It is the first significant merger since 2000, and the first of many to follow in the wave of industry consolidation. US Airways, the nation's then seventh-largest airline, and America West

[1] See, for example, Borenstein and Rose (1994), Stavins (2001), Geradi and Shapiro (2009) and Dai, Liu and Serfes (2014).

(the eighth-largest), were reported to have 361 planes, with 44,100 employees and $10 billion in annual revenue at the time of merger. They had little overlap in the routes they operated on: US Airways concentrated in the east coast while America West in the west coast. Their merger did not immediately change how the two companies operated, but they were to coordinate their schedules and integrate their frequent flier programs over time. All flights would operate under US Airways' name after merger was completed.

For on-time performance, we focus on arrival delay and we find that merger reduces the merging carrier's arrival delay by about 1.3 minutes.[2] Treating arrival delay as a measure of quality (vertical differentiation), we then extend the analysis to other measures of vertical differentiation. These measures include number of flights per route, number of routes and number of destinations. Using matching method, we find significant impact of merger on all these measures. In particular, after merger, number of routes goes down by about 116 and number of destinations decreases by about 1.96. Number of flights goes up after merger, by almost 14 flights per route. Ignoring price impact, merger benefits consumers with lower arrival delay and more flights, but hurts consumers with fewer options (routes and destinations) to choose from.

Next, we analyze another aspect of differentiation across airlines - horizontal differentiation. In contrast to vertical differentiation where all consumers prefer high quality to low quality (everything else the same), under horizontal differentiation, some consumers prefer one product while others prefer another. Horizontal differentiation in the airline industry has scarcely been analyzed. We model and measure horizontal differentiation by comparing the scheduled departure times

[2] To put this into perspective, average arrival delay in our sample is 12.277 minutes. Mazzeo (2003) finds that a flight on a monopoly route is on average 1.35 minutes later than a similar flight on a competitive route.

(view scheduled departure time as a characteristic of the products). If flights have the same scheduled departure time, then there is minimum product differentiation on this product characteristic. But if one flight departs at 8am while another at 10am, then different travelers may have different preferences between these two flights (horizontal differentiation). To measure the degree of horizontal differentiation, for each flight, we calculate the gap (in minutes) between this flight and the closest (in terms of departure time) other flight. In one setting (*All* carriers), we impose no constraint on the identity of the carrier of the closest flight. In the other setting (*Between* carriers), we restrict the closest flight to be from a different carrier. Using the gaps for flights, we then construct two measures of horizontal differentiation: one scale-dependent (*Gap*) and the other scale-independent (*Gini*).[3] Our results show that merger reduces *Gap_All* (gap in the *All* carriers setting), but has only insignificant impact if we restrict the identity of next flight's carrier to be from different carriers (*Gap_Between*) or if we consider scale-independent measures (*Gini_All* and *Gini_Between*). It is possible that this merger has little impact on the degree horizontal differentiation, or perhaps, or more likely, that the impact was not captured by our specific measures of horizontal differentiation. More research is needed in this direction to improve our general understanding of merger impact.

1.1.1 Literature review

Our paper is closely related to the literature studying on-time performance in the airline industry, in particular the link between on-time performance and competition. There have been increasing concerns regarding both increasing market

[3]More details about the construction of these horizontal differentiation measures can be found in the Appendix.

concentration and worsening flight delays in the U.S. airline industry. Rupp, Owens and Plumly (2001) find that when routes become more competitive, on-time performance gets worse. In contrast, Mazzeo (2003) finds that relative to routes with competition, monopoly routes experience significantly more and longer flight delays. In our paper, we use panel data together with matching method to analyze the impact of merger on product differentiation, with the commonly studied arrival delay as one measure of product differentiation.

Our paper is also related to the literature studying the impacts of merger. Much of the focus is on how merger affects prices which typically centers on two opposite mechanisms (see, for example, Carlton et. al. (1980), Kim and Singal (1993) and Focarelli and Panetta (2003)). On one hand, merging firms may take advantage of economies of scale or economies of scope, leading to lower costs and potentially lower prices (efficiency effect). On the other hand, merger increases market concentration and higher prices can be sustained as a result (market power effect). Overall, the impact of merger on prices can be ambiguous. This ambiguity in theory carries through to empirical settings where the findings are also mixed, using data from the airline industry, the banking sector as well as the hospital industry (e.g., Kim and Singal (1993), Prager and Hannan (1998) and Dafny (2005)).[4] In contrast to the abundance of literature on the price impacts of merger, little has been done on the impact of merger on quality, at least for the airline industry.[5] The mechanisms of quality impacts are in similar spirit as the mechanisms of price impacts. For example, merger allows merging carriers to combine their resources and consolidate their services, potentially leading to higher qualities. On the other hand, with less competition, merging carriers may feel less pressure to compete on quality and thus may have an incentive to reduce

[4]See Prince and Simon (forthcoming) for a more detailed discussion of this literature.
[5]Exceptions include Prince and Simon which we will discuss in detail next.

their investment in quality.[6]

Our paper is most closely related to Prince and Simon who analyzes the impact of merger on on-time performance in the U.S. airline industry. Prince and Simon analyze five recent U.S. airline mergers, focusing on several measures of on-time performance. Distinguishing short-run and long-run impacts of merger, they find that merger tends to reduce merging carrier's on-time performance in the short run (maybe due to the challenge of integrating the merging carriers), but improves it in the long-run due to merger-induced efficiency gains. Our paper differs from Prince and Simon in several perspectives. First, we consider the more general question of product differentiation and analyze how it is impacted by merger.[7] We consider both vertical and horizontal differentiation, and on-time performance is only one of the measures for vertical differentiation. Second, we focus on a specific merger (that between US Airways and America West in 2005) and carefully select sample periods to isolate the impact of merger and avoid the contamination of other events (e.g., 9/11, financial crisis, other mergers).[8] Third, and probably most importantly, while we all use diff-in-diff method to identify the impact of merger, we used different comparison group to compare with the treated group (i.e., the merging carriers and their routes). The initial control group contain all routes which only non-merging carriers operate on.[9] We then use matching method to identify carrier-routes from the control group that

[6]In the case of airline industry, price effect has a perverse impact on quality. That is, by raising price, demand and supply goes down which reduces the congestion of air travel, and leads to better on-time performance.

[7]There are earlier studies looking at product differentiation in the U.S. airline industry but not how it is impacted by merger. See, for example, Berry (1990) and Borenstein and Netz (1999).

[8]The downside of this approach is that since we only include the limited time after the merger (before another merger takes place), our results are only for the short to medium-run impacts.

[9]We use control group and comparison group interchangeably. We do not include any of the routes which merging carriers also operated on pre-merger, i.e., the overlapping routes. We expect merger to affect the non-merging carriers on these routes as well. Correspondingly our results are silent on the impact of merger on the overlapping routes.

mimic the carrier-routes in the treated group. These selected carrier-routes from the control group then form the matched group. Next, we compare the treated group with the matched group to isolate the impact of merger.

1.2 Econometric Model

In this section we introduce our econometric model and discuss the identification strategy to estimate the impact of airlines merger on various measures of vertical and horizontal product differentiations. First, we set up the model and discuss why the standard estimation method such as fixed effect model and differences-in-differences (DID) method provide inconsistent estimates of the impact of merger on different measures. In addition, we discuss how propensity score matching addresses the potential self selection of merging airlines and the limitations of this method applying in our panel data. Finally we introduce the robust genetic matching method to address the potential endogeneity of the impact of two airlines merger on various outcome variables.

1.2.1 Model Setup

We define treatment as two airlines merge together. Thus the treated airlines are US Airways and America West and the rest of the other seven airlines belong to the control group. To estimate the impact of merger between US Airways and America West on different vertical and horizontal product differentiations, we consider the following fixed effect model:

$$Y_{ijt} = \alpha + \mu_{ijt} + \gamma_{ijt} + \beta_1 \mathrm{Trt}_{ijt} + \beta_2 X_{ijt} + \epsilon_{ijt} \tag{1}$$

where the outcome variable Y_{ijt} represents various measurements of vertical and horizontal product differentiation and the subscripts i, j, t denote carrier, route and time respectively. As discussed the variable of interest Trt_{ijt} is a binary indicator for treatment carriers in the post-merger periods. The set of control variables X_{ijt} includes a vector of route and carrier level characteristics.[10] In addition, we include year month fixed effects (μ_{ijt}) and carrier route fixed effects (γ_{ijt}).

The decision of merger is potentially endogenous because whether two airlines merge or not depends on their specific characteristics as well as the characteristics of the operating routes. Merging and non-merging airlines differ over many aspects such as financial conditions, route structure, market shares and many others. The average treat effect estimator β_1 of the above fixed effect model is biased and inconsistent because the unobserved heterogeneity between merging and non-merging carriers vary over time and this fixed effect model can only control the time invariant unobserved hetrogeneity.

Next we consider the following differences-in-differences model:

$$Y_{ijt} = \alpha + \beta_1 \text{Trt Carrier}_{ij} + \beta_2 \text{Post Trt}_t + \beta_3 \Big(\text{Trt Carrier}_i \times \text{Post Trt}_t\Big) + \beta_4 X_{ijt} + \epsilon_{ijt} \tag{2}$$

where Trt Carrier_{ij} is a binary variable for treatment carrier and Post Trt_t is the post-merger time dummy. The coefficient of the interaction term $\text{Trt Carrier}_{ij} \times \text{Trt Merger}_t$, β_3 captures the average treatment effects of merger between US Airways and America West on various outcome measures.

The coefficient of interest β_3 in equation (2) is consistent under the assumptions that merger is exogenous and both treated and control group airlines have similar trend in pre-merger time period. Clearly, two airlines self selected for

[10]See the summary statistics table for details about the controls.

merger and hence we can not treat this event as exogenous. Since treated and control group carriers differ on their observable and unobservable characteristics, it is highly unlikely that they share similar trends in terms of the different outcome variables. Therefore, DID estimates are also biased and inconsistent.

1.2.2 Propensity Score Matching

To address the potential self-selection of merger between US Airways and America West, we adopt the propensity score matching because according to Rubin(1973), Heckman, LaLonde and Smith (1999) this method can largely reduce the bias from endogenous selection conditional on the balanced observable covariates. Another advantage of this method is that it selects most similar comparison groups in terms of covariates, and the difference between treatment group and matched comparison group is the treatment status. Thus, in a quasi-experimental study the average treatment effect on treatment can be drawn by comparing the matched treatment and control groups.

The outcome variable Y_{ij} denotes various measure of vertical and horizontal and product differentiations and is defined as

$$Y_{ij} = \begin{cases} Y_{ij}(1) \quad if \quad \mathrm{Trt}_{ij} = 1 \;\; \textit{that is } i \in \{ \text{ US Airways}, \;\; \text{America West}\}, \\ \\ Y_{ij}(0) \quad if \quad \mathrm{Trt}_{ij} = 0 \;\; \textit{that is } i = \;\; \text{Any other airlines} \end{cases}$$

Following Rosenbaum and Rubin (1983) the average treatment effect on the treated (ATT) airlines US Airways and America West can be written as

$$\gamma = \mathbb{E}\Big[Y_{ij}(1) - Y_{ij}(0) \mid \mathrm{Trt}_{ij} = 1\Big]$$

This expression cannot be estimated directly, because $Y_{ij}(0)$ is not observed for treated airlines. To estimate the ATT(Average Treatment on the Treated) we use the conditional independence and common support assumptions[11].

The conditional independence assumption is the key identifying assumption of propensity score matching estimation. Since this assumption is fundamentally un-testable, we use a rich set of covariates including carrier level, route level and carrier-route specific characteristics to meet this assumption. Under this assumption, the average treatment effect for the subpopulation with $X_{ij} = x_{ij}$ equals

$$\begin{aligned}\gamma &= \mathbb{E}\Big[Y_{ij}(1) - Y_{ij}(0)|X_{ij} = x_{ij}\Big] \\ &= \mathbb{E}\Big[Y_{ij}|Trt_{ij} = 1, X_{ij} = x_{ij}] - \mathbb{E}[Y_{ij}|\mathrm{Trt}_{ij} = 0, X_{ij} = x_{ij}\Big]\end{aligned} \tag{3}$$

Heckman et al. (1998) and Smith and Todd (2004) argue that the propensity score matching method minimizes the discrepancies of the observable characteristics between treated group and control group when the sample selection is due to observable characteristics. Thus by using propensity score matching we find a reliable counterfactual control group since we have a very rich set of control variables.

[11]The conditional Independence means the treatment status Trt_{ij} condition on X_{ij} is as good as random that is selection on observable covariates, X_i we obtain $Y_{ij}(1), Y_{ij}(0) \perp Trt_{ij}|X_{ij}$. The common support assumption implies that for all carrier i in treatment group, there is a positive probability of either merging ($Trt_i = 1$) or not merging ($Trt_i = 0$), that is $\eta < Pr(Trt_{ij} = 1|X_{ij} = x_{ij}) < 1 - \eta$ for some $\eta > 0$.

1.2.3 Mahalanobis Distance

In propensity score matching method we choose a control group based on the similar values of the unidimensional distance metric propensity score.[12] Estimated propensity scores are one distance metric among many alternatives. Further, the preference for a logit or probit specification is rarely justified in any formal sense rather is typically adopted for simplicity. Another common method of multivariate matching is based on Mahalanobis distance (CoChran and Rubin (1973), Rubin (1979, 1980)). The Mahalanobis distance metric between units with covariates x_{ijt} and x'_{ijt} is defined as

$$D_M = \sqrt{\big(x_{ijt} - x'_{ijt}\big) S \big(x_{ijt} - x'_{ijt}\big)'} \tag{4}$$

where S is the sample covariance matrix. Therefore, Mahalanobis distance metric measures the multivariate distance between individuals in different groups.

Rubin (2001) and Rosenbaum and Rubin (1985) show that Mahalanobis distance and propensity score matching can be compared in various ways. Under normality assumption both the matching staisfy the equal percentage bias reduction property.[13] Further, the resulting set of matches is invariant to affine transformations of the covariates for both types of matching. However, Mahalanobis distance matching does not perform well when covariates have nonellipsoidal distributions. Because of this limitation of Mahalanobis distance matching, Rosenbaum and Rubin (1983) suggested matching on propensity score instead. However, due to sampling variation and nonexact matching $Trt \perp X|P(X)$ may

[12]The propensity score $P(X_i)$ is the conditional probability of assignment to treatment given the covariates: $P(X_i) = Pr(Trt = 1|X_i) = E(Trt_i|X_i)$.

[13]This property implies that the percentage bias reduction is the same for all linear combinations of regression coefficient β.

not hold after matching on the propensity score. Therefore, Rosenbaum and Rubin (1985) argue that in addition to propensity score, one should match on individual covariates by minimizing the Mahalanobis distance to have balance covariates. Hence, Rosenbaum and Rubin (1985) recommend that one should first match on the propensity score and then match based on Mahalanobis distance within propensity score strata.

1.2.4 Genetic Matching

The genetic matching is based on the GenMatch algorithm proposed by Diamond and Sekhon (2013). This algorithm searches a range of distance metrices to find the particular measure that optimizes post-matching covariate balance. In particular, the GenMatch algorithm is most applicable when the Mahalanobis distance is not optimal for achieving balance in a given dataset. GenMatch algorithm assigns a weight W_i to each potential distance metric for all matching variables. In addition, the algorithm also weights each variable according to its relative importance for achieving the best overall balance. To implement the GenMatch algorithm we need to specify a specific loss function. Diamond amd Sekhon (2013) propose a generalised version of Mahalanobis distance which include an additional weight matrix:

$$GD_M = \sqrt{\left(x_{ijt} - x'_{ijt}\right)\left[S^{-1/2}W(S^{-1/2})'\right]\left(x_{ijt} - x'_{ijt}\right)'} \tag{5}$$

where W is a $k \times k$ positive definite weight matrix and $S^{-1/2}$ is the Cholesky decomposition of S. All elements of W are restricted to 0 except those down the main diagonal which consists of k parameters that must be chosen.

Diamond and Sekhon (2013) recommend to include the propensity score as one of the covariates. Thus X in equation (5) is replaced by Z where $Z = [P(X)\ X]$. If optimal balance is achieved by simply matching on the propensity score, then a zero weight will be assigned to all other covariates and genetic matching will be equivalent to propensity score matching. Alternatively, when we assign zero weight to the propensity score and weight 1 to every other variable in Z, genetic matching becomes numerically equivalent to Mahalanobis distance matching. Therefore, both propensity score and Mahalanobis distance matching can be considered special cases of genetic matching. However, in general the GenMatch algorithm finds that neither minimizing the MD nor matching on the propensity score minimizes the loss function and thus searches for improved metrics that optimizes covariate balance.

Diamond and Sekhon (2013) show that searching all possible weights to optimize the covriate balance for any given combinations of covariates, GenMatch algorithm improves the overall covariate balance and gurantees asymptotic convergence to the optimal matched sample. Therefore, by construction the algorithm improves covariate balance for a specific loss function chosen by the applied researchers. we use a loss function that minimizes the overall imbalance by minimizing the largest individual discrepancy, based on $p-$values from KS tests and paired t-tests for all variables that are being matched on. This loss function also includes individual balance measures that are sensitive to many forms of imbalance, such as KS test statistics, and not simply difference of means tests. Therefore, the GenMatch algorithm finds the weight W which optimizes the above specified loss function. Diamond and Sekhon (2013) show that increasing population size usually improves the overall balance achieved by this algorithm. Since we use a large sample, GenMatch algorithm give us a robust control group which satisfies the overall covariates balance.

To make sure that our matched control group has a similar time trend as the treatment group in the pre-merger time period we include a set of lag dependent variables in Z while applying the GenMatch algorithm and to estimate the average treatment effect on the treated (ATT), we use the differences-in-differences (DID) matching method developed by Heckman, Ichimura and Todd (1997) and Heckman, Ichimura, Smith and Todd (1998). This estimator is analogous to the standard DID regression estimator but we use a selected matched sample as a control group instead of the entire comparison group. By construction our matched control group is more similar to the treatment group.

We obtain the matched control group using the GenMatch algorithm along with its generalized distance metric as shown in equation (5). However, the GenMatch algorithm does not guarantee that the matched control group always have a similar time trend. Therefore, we perform a test to check that whether the common time trend assumption holds or not for each different outcome variables. The test is based on the linear regression model:

$$Y_{ijt} = \delta_0 + \delta_1 t + \delta_2 \big(t \times \text{Treatment}_i\big) + X\beta + \eta_{ijt} \tag{6}$$

where t represents the linear time trend, Treatment_i is the dummy variable for merger airlines and X is the set of control variables. The statistical significance of the coefficient δ_2 implies that treatment and control groups have different linear time trend. Note that if the common time trend assumption does not hold, the coefficient of interest β_3 in equation (2) is biased.

1.3 Application to US Airways – America West merger

In this section, we apply the matching method to study the impact of the US Airways – America West merger. We first discuss the data and present some preliminary estimation results.

1.3.1 Data and preliminary estimation results

To analyze the impact of merger between US Airways and America West, we use four sources of data, three of which from the Bureau of Transportation Statistics (BTS). First is the On-Time Performance (hereafter OTP) data which provides scheduled departure/arrival time, departure and arrival delays (in minutes) for every non-stop domestic flight. Another data set we use is Air Carrier Financial Reports, which provide financial information for the U.S. airlines. The third data from BTS is Airline Origin and Destination Survey (DB1B), which allows us to construct market structure variables such as HHI and market share. We also use the MSA population data from Census.

Let us start with the OTP data which covers US certified airlines that account for at least one percent of domestic scheduled passenger revenues. This dataset provides information about on-time performance for non-stop domestic flights. Merger between US Airways and America West was proposed on May 19, 2005 and approved on September 27, 2005. Our sample period starts on January 1, 2003 and ends on September 30, 2008. We choose this time period to minimize the impact of confounding factors unrelated to the US Airways and America West merger. In particular, we choose January 2003 as the starting period to minimize the impact of: (i) acquisition of TWA by American Airlines (in April 2001) and (ii) 9/11 terrorism and the ensuing (demand and supply) shock on the airline industry. Delta and Northwest announced their merger in April 2008 which was approved by the Department of Justice in October 2008. To avoid the contamination of this subsequent merger, our sample does not include the last quarter of 2008 or later. Since on-time performance is observed for each directional flight, we define a route as a directional pair of origin and destination airports, as is common in studies on airline on-time performance. For example, ATL-DFW and DFW-ATL are two different routes which have different on time performance. While on-time performance data (e.g., arrival delay) is at the flight level, we aggregate data typically into carrier-route-quarter cells for our empirical analysis. Other measures of vertical differentiation are calculated similarly. The measures of horizontal differentiation are more tedious to derive and we discuss them in detail in the Appendix.

Air Carrier Financial Reports are at the quarterly level, and cover large certificated U.S. carriers with annual operating revenue of $20 million or more. It is a carrier level dataset, and we generate financial variables such as current liability and current assets at the carrier-year level from this data source. Airline Origin and Destination Survey (DB1B) contains summary ticket level data, which is a 10% sample of airline tickets from reporting carriers. We use DB1B data to construct market structure variables such as number of carriers, market share and HHI.[14]

We first report the summary statistics (at carrier-quarter level) which are provided in Table 1.1. For quality measures, the average arrival delay is about 12.277 minutes. On average, each carrier operates on about 451 routes in a quarter, and offers about 289 flights per route in a month. We also calculate

[14]Market structure variables can also be constructed using OTP data. However, only carriers accounting for at least one percent of domestic scheduled passenger revenues are required to report for OTP data. As a result, OTP data will systematically overlook smaller carriers, leading to less accurate description of market structure.

number of destinations at the carrier-route-quarter level, defined as geometric mean of the number of destinations for each of the two end airports for a carrier-route-month. Average number of destination is about 12.8. Note that all these stats are calculated with data both before and after-merger. Table 1.1 also reports the measures of horizontal differentiation. For example, the average gap between a flight and its closest other flight is about 10 minutes.[15]

We consider the vertical differentiation (or quality) measures. On-time performance is clearly a quality measure and we focus on arrival delay. In addition, we consider number of flights, number of routes and number of destinations as quality measures as well. More flights on a route give travelers more options to choose from, and more likely to find a flight closer to their ideal one. More routes and destinations give a traveler more options to fly with a single carrier, thereby accruing frequent flier miles more quickly. It also gives the traveler more choices when redeeming their frequent-flyer miles.

For horizontal differentiation, we construct and use the following measures: *Gap_All*, *Gini_All*, *Gap_Between* and *Gini_Between*. They are derived using flight scheduled departure times.[16] Consider two flights with scheduled departure times at 9am and 10am respectively. Some travelers may prefer the 9am flight while others may prefer the 10am flight. Hence this is horizontal differentiation as opposed to vertical differentiation where everyone prefers a product to another.[17]

[15] When calculating *Gap_Between*, we restrict that the closest flight has to be from a different carrier. This in general should increase the gap (in minutes). Therefore, we should expect the mean of *Gap_All* to be smaller than the mean of *Gap_Between*. However, the restriction of a different carrier requires at least two carriers in the market, i.e., we have to drop all monopoly routes. Similarly, when calculating *Gini_Between*, there have to be at least 3 flights to calculate the dispersion of Gap. Dropping out the monopoly routes when calculating *Gap_Between* likely is responsible for the counterintuitive stat that mean of *Gap_All* is larger than *Gap_Between*.

[16] See the Appendix for more details on how these measures are constructed.

[17] Flight scheduled departure times may also have a vertical flavor. For example, most travelers are likely to prefer a 10am flight to a 5am flight or an overnight flight. However, there is a key difference between our vertical and horizontal differentiation measures. Our vertical

Table 1.1: Summary Statistics for Airline Merger

Outcome					Control				
Variable	**Sample**	**Treatment**	**Control**	**Diff**	**Variable**	**Sample**	**Treatment**	**Control**	**Diff**
Arrival Delay (ijt)	12.277	11.003	12.406	1.403	HHI (jt)	.679	.604	.687	.083
	(10.646)	(8.925)	(10.797)	(13.026)		(.259)	(.233)	(.261)	(31.577)
Num Flights(ijt)	288.691	292.781	288.277	-4.504	Mkt Share (ijt)	.628	.536	.637	.101
	(242.103)	(232.718)	(243.029)	(-1.838)		(.337)	(.316)	(.338)	(29.829)
Num Dest(ijt)	12.814	14.537	12.639	-1.897	Num Carriers (jt)	2.785	3.273	2.736	-.537
	(7.254)	(7.822)	(7.171)	(-25.914)		(1.790)	(1.763)	(1.786)	(-29.748)
Num Routes (it)	450.565	478.802	447.708	-31.094	Cancelled Flights (ijt)	5.287	4.594	5.357	.763
	(179.558)	(51.802)	(187.463)	(-17.130)		(9.416)	(8.866)	(9.468)	(8.011)
All Gap (jt)	12.051	11.637	12.095	.458	Distance (jt)	886.835	1005.195	874.861	-130.333
	(15.604)	(16.607)	(15.493)	(2.589)		(656.264)	(715.433)	(648.781)	(-19.654)
All Gini (jt)	.263	.295	.260	-.035	Passengers (ijt)	1026.300	1053.688	1023.529	-30.158
	(.143)	(.139)	(.143)	(-21.422)		(1385.188)	(1085.130)	(1411.970)	(-2.151)
Bet Gap (jt)	5.413	5.456	5.408	-.048	Load Factor (ijt)	.726	.725	.726	.001
	(10.467)	(11.917)	(10.286)	(-0.332)		(.127)	(.131)	(.127)	(1.098)
Bet Gini (jt)	.316	.339	.314	-.025	Mkt Fare (ijt)	196.571	198.438	196.383	-2.055
	(.137)	.127	.138	-13.381		(77.853)	68.682	78.719	-2.608

For each measurements, the first row is the mean value of the variable, second row is the t test statistics and last row is the number of observations. carrier: i, route: j, month: t

Next, we investigate the impact of merger using estimation methods. We will start with OLS and simple difference-in-difference (thereafter diff-in-diff). The results are provided in Table 1.5. For OLS, our additional controls include HHI, market share, number of carriers on the route, passengers, ticket price, load factor and scheduled departure flights.

Diff-in-diff and the construction of "control" group

With merger (treatment) affecting some observations but not others, diff-in-diff seems a natural choice to tease out the common trend which is not due to merger and isolate the impact of merger. The treated group includes all routes operated by the merging carrier(s). The comparison group consists of all other carriers, but only routes where the merging carriers do not operate on. That is, we only use non-overlapping routes to form the comparison group. This is because, merger will affect the merging carriers on the routes they operate on. This in turn affects the non-merging carrier on the overlapping routes, making them inappropriate as comparison groups to study the impact of merger. In the meantime, U.S. airline industry consists of only few major airlines (especially so after the recent merger wave), with each operating on a significant portion of the routes. Removing all overlapping routes thus significantly limits the choice and size of comparison group, making it more difficult for diff-in-diff method to work well.[18] The additional control variables (estimates not reported) are the same as in OLS regression. Also, note that for diff-in-diff to work well, the treated group

differentiation measures vary across carriers as it measures each carrier's (average) product quality. Our horizontal measures is common across carriers as they are aggregations of how the products by different carriers differ from each other in a market.

[18]This problem is less severe here since the two merging airlines are both ranked outside top 5 at the time of merger. However, in our falsification tests later on, we will introduce fake mergers, for example, between American Airlines and Alaska Airlines. To construct, we exclude all routes which the true merging carriers operate on (US and HP), as well as the routes which the fake merging carriers (AA and AS) operate on. The latter further limits the choice and size of comparison group.

and the "control" group need to have the same common trend.

With simple diff-in-diff, we use the whole comparison group to control the common trend. The problem is that merger is not an exogenous event – the fact that US Airways and America West are merging but not other airlines suggest that there is something special about the merging carriers.[19] Controlling carrier and route characteristics helps reduce this selection/endogeneity issue, but one can do better. And this is what we we aim to do with diff-in-diff matching. The main idea is to select a subsample out of the whole comparison group which best resemble the "treated group", to form a matched group. We then run diff-in-diff using the treated and matched group, rather than using treated and whole comparison group. Matched group better resembles the treated group because that's the criteria based on which they are chosen, and we will present a series of evidence. But first we need to explain how the matching methodology works.

1.3.2 Market characteristics used for matching

We first discuss the variables we use for matching. There is no uniform theory on what characteristics are good predictors for the various measures of vertical and horizontal differentiation. Much of the empirical literature has shown that market structure is important in capturing market competition and the corresponding outcomes (e.g., price level, price dispersion). In line with this literature, we use several market structure variables including market share and number of passengers (at the carrier-route-month level), and number of carriers and HHI (at the route-month level). Both are then aggregated from month to quarter-level. We also use ticket price (at the carrier-route-quarter level) to take into account

[19] This is confirmed later on – the common trend assumption is rejected for several measures of vertical differentiation.

the cost/income heterogeneity across markets. Some route-level characteristic are not particular to the airline industry, for example, distance. One variable of particular interest to us is arrival delay which has been analyzed in various existing studies. Other variables we choose include the number of canceled flights and load factor, all aggregated to carrier-route-quarter level. We also want to use historical arrival delays as part of the variables for matching. Combined, we include all the explanatory variables in the last pre-merger period, as well as all lags of the arrival delay.

One may think that carrier level variables (e.g., financial variables, bankruptcy status) are also important. We did not include them for two reasons. First, carrier characteristics is already partially reflected in the market structure variables (e.g., market share). Second, we want to reserve these carrier-level variables as part of the control variables when we run diff-in-diff estimation after matching is done.

1.3.3 "Control" group: Matched vs. Comparison group

Using the matching method described in the previous section, we can construct matched group(s) for each dependent variable.[20] Since the criteria for a carrier-route in the comparison group to be selected into the matched group is that they best resemble the treated group, we should expect the matched group to better resemble the treated group than the whole comparison group. In an effort to establish support for this, next, we present a series of comparisons/tests.

[20]Briefly reiterate why we have different matched groups – this is because we use lagged dependent variable in the matching process.

Summary stats comparison

We first compare the summary stats for treated group, matched group(s) and whole comparison group. Recall that matching is mostly based on a series of explanatory variables, plus the lagged values of the dependent variable (e.g., arrival delay). Summary stats for these variables are included in Table 1.2. We can say that relative to the comparison group, matched group has mean values closer to that of the treated group, for example, the market structure variables (market share, HHI, number of carriers) and route characteristics (distance, passengers).[21] Moreover, the matched group typically has lower standard errors as well.

Next, rely on more rigorous tests.

Common trend tests comparison

Table 1.3 presents the results for common trend tests. We can see that if the whole comparison group is used, then the common trend assumption fails for all 4 vertical differentiation measures. If we use matched group(s), they still fail for two measures, but now passes for number of destination, and passes for arrival delay as well if we use all 9 lags of arrival delay in the matching process. One can also find that matched group lead to lower t-stats on average. These all seem to suggest that using matching to study merger impact potentially improves the accuracy relative to standard diff-in-diff.[22]

[21] An exception is canceled flights for which the comparison group better resemble the treated group.

[22] Of course, much more work is needed as common trend test still fails for several measures even when using the matched group.

Table 1.2: Summary Statistics All Pre-merger Periods for Airline Merger

Control only for Arrival Delay			
Variable	**Treatment**	**Comparison**	**Matched**
HHI (jt)	.63	.685	.617
	(.236)	(.252)	(.221)
Mkt Share (ijt)	.603	.688	.615
	(.3)	(.291)	(.272)
Num Carriers (jt)	3.335	2.932	3.134
	(1.932)	(1.89)	(1.827)
Cancelled Flights (ijt)	5.416	5.207	4.409
	(9.536)	(9.848)	(7.707)
Distance (jt)	1002.289	893.944	982.373
	(679.861)	(639.133)	(642.287)
Passengers (ijt)	1217.621	1327.405	1194.238
	(1158.178)	(1533.834)	(1089.138)
Load Factor (ijt)	.717	.706	.725
	(.121)	(.118)	(.112)
Mkt Fare (ijt)	195.243	189.077	200.579
	(60.598)	(77.096)	(69.215)
Arrival Delay (ijt)	10.256	10.152	10.031
	(5.068)	(7.064)	(4.537)

For each measurements, the first row is the mean value of the variable, second row is the t test statistics and last row is the number of observations. carrier: i, route: j, month: t

Table 1.3: Common Trend Comparison: Simple DiD vs. DiD Matching(9 lags) vs. DiD Matching(4 lags)

Variable	(1) Simple Diff-in-Diff	(2) DiD Matching(9 Lags)	(3) DiD Matching(4 Lags)
Arrival Delay	-0.067*	0.004	-0.129***
	(-2.51)	(0.08)	(-3.36)
Num Flights	3.967***	0.836	0.726
	(4.07)	(1.58)	(1.31)
Num Dest	0.380***	0.338***	0.276***
	(10.46)	(8.73)	(6.66)
Num Route	10.590***	10.090***	12.740*
	(4.63)	(6.70)	(2.45)
All Gap	-0.059	-0.073	0.042
	(-0.76)	(-1.00)	(0.51)
All Gini	0.001	-0.000	0.000
	(0.68)	(-0.18)	(0.15)
Bet Gap	-0.015	-0.028	-0.000
	(-0.26)	(-0.58)	(-0.01)
Bet Gini	-0.003	-0.003	-0.003
	(-1.96)	(-1.59)	(-1.88)

For each measurements, the first row is the coefficient for the variable in interest, second row is the t test statistics. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Falsification tests comparison

Next, we conduct two sets of falsification tests. In the first set, we include only the pre-merger data which includes 10 quarters of data. We assign a fake treatment date at the middle, with the first 5 quarters as the *pre-merger* period and the last 5 quarters as the "post-merger period." We then construct matched group and run diff-in-diff as in matching diff-in-diff for the main model. The results are reported in Table 2.6, third panel of **Fake Merger Time**. We report only the estimate of the interaction term, which measures the impact of this fake "merger" on the "treated group" on various product differentiation measures. We see that the fake merger has no impact on most horizontal measures of product differentiation, but seems to have impacts (significantly different from zero) on some vertical measures. Moreover, matched group in general does better than the whole comparison group. For example, when Between_Gini is dependent variable, our matching diff-in-diff pass the falsification tests but the standard diff-in-diff fails (at 5% level). Also, the estimates for matched groups are in general smaller (in magnitude) and have smaller t values relative to the estimates for the whole comparison group. Of course, the matched groups still fail some falsification tests, suggesting that effort to construct a better matched group is still much needed.

In the second set of falsification tests, we include both pre- and post-merger data, but assigning two non-merging carriers to a fake merger. We sort airlines alphabetically by their two-digit codes. Two fake mergers were considered. In **AAAS**, we pick the first two airlines: AA and AS, or American and Alaska airlines, to "merge". In **UAWN**, the last two airlines are assigned to the "treated" group: UA and WN (United and Southwest). For each fake merger, the comparison group includes all routes that is not operated by airlines involved in either fake merger or the actual merger. We then use similar matching method to construct matched group. Their diff-in-diff results are presented in Table 2.6 columns 2 and 5.[23] We can see that matched group again has smaller magnitude of t values, and the difference sometimes is significant in that the matching diff-in-diff passes the falsification tests when the standard diff-in-dif fails. For example, see *All_Gap* and *Between_Gini* for **AAAS** fake merger, and *Arrival delay* and *All_Gini* for **UAWN** fake merger.

diff-in-diff results comparison

We have shown that the whole comparison group and the matched group differ in their composition, summary stats, common trend tests and falsification tests. A natural question then is, if we use them to run diff-in-diff, would they still lead to similar estimates? Table 1.5 present the estimation results which suggest that the answer is no.

We will use the 9-lag matching to compare with the simple diff-in-diff. Simple diff-in-diff results show that merger has significant impact on two horizontal differentiation measures: All_Gini and Between_Gini, but the matching diff-in-diff results show smaller (in magnitude) and insignificant impacts. For all vertical measures, both set of results show significant merger impacts, but their magnitude can be quite different. Diff-in-diff matching results show that merger reduces number of destinations by about 2 and a far more drastic reduction of 116 for number of routes. This only leads to a mild increase of 2 flights per route after merger. Merger also reduces arrival delay about 1.3 minutes. These merger impacts are larger than their counterparts in column (2) by 40% or more except for number of destinations.

[23] Again we report estimates for the interaction term only which measures "merger" impact.

Table 1.4: Falsification Test for Airline Merger

	AAAS			UAWN			Random Time		
Variable	(1) Diff-in-Diff	(2) DiD Matching (9 Lags)	(3) DiD Matching (4 Lags)	(4) Diff-in-Diff	(5) DiD Matching (9 Lags)	(6) DiD Matching (4 Lags)	(7) Diff-in-Diff	(8) DiD Matching (9 Lags)	(9) DiD Matching (4 Lags)
Arrival Delay	0.597***	0.708**	0.657**	-0.396**	-0.063	-0.149	-0.270	0.147	0.177
	(3.83)	(2.97)	(2.70)	(-3.16)	(-0.32)	(-0.74)	(-1.81)	(0.71)	(0.86)
Num Flights	-7.465	1.345	3.089	-9.674	-2.748	-8.054	4.065	3.948	4.419
	(-1.46)	(0.48)	(0.82)	(-1.52)	(-0.72)	(-0.79)	(0.93)	(1.84)	(1.78)
Num Dest	-2.078***	-1.083***	-1.593***	1.073***	1.536***	1.556***	1.870***	1.576***	1.578***
	(-12.93)	(-4.68)	(-6.83)	(8.40)	(8.63)	(6.26)	(14.00)	(9.89)	(9.01)
Num Route	-89.26***	-47.73*	-50.62*	62.69*	82.10**	82.10**	61.02***	48.39***	69.72***
	(-4.15)	(-3.05)	(-2.61)	(2.69)	(6.36)	(6.36)	(83.13)	(52.07)	(22.42)
All Gap	0.919***	0.118	0.227	0.056	-0.029	-0.183	-0.127	-0.181	0.053
	(3.87)	(0.41)	(0.70)	(0.22)	(-0.08)	(-0.49)	(-0.44)	(-0.58)	(0.15)
All Gini	-0.043***	-0.024***	-0.025***	-0.019***	-0.011*	-0.007	-0.000	-0.004	-0.002
	(-8.51)	(3.45)	(-3.75)	(1.73)	(-1.99)	(-1.21)	(-0.04)	(-0.75)	(-0.27)
Bet Gap	0.307	0.345**	0.172	-0.094	-0.174	0.380*	0.026	-0.177	-0.060
	(1.84)	(2.82)	(1.18)	(-0.70)	(-0.97)	(2.23)	(0.12)	(-0.77)	(-0.25)
Bet Gini	-0.032***	-0.015	-0.023*	0.000	-0.011	0.009	-0.014*	-0.014	-0.014
	(-5.46)	(-1.56)	(-2.21)	(0.03)	(-1.22)	(0.78)	(-2.19)	(-1.54)	(-1.52)

For each measurements, the first row is the coefficient for the variable in interest, second row is the t test statistics. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Results Comparison: OLS vs. Simple DiD vs. DiD Matching(9 lags) vs. DiD Matching(4 lags)

Variable	(1) OLS	(2) Simple Diff-in-Diff	(3) DiD Matching (9 Lags)	(4) DiD Matching (4 Lags)
Arrival Delay	-0.232	-0.925***	-1.313***	-1.075***
	(-1.65)	(-6.04)	(-5.27)	(-3.76)
Num Flights	1.159	7.773*	13.78***	11.26***
	(0.30)	(2.12)	(4.21)	(4.05)
Num Dest	-2.523***	-2.033***	-1.964***	-2.095***
	(-15.05)	(-12.09)	(-9.40)	(-8.51)
Num Route	-97.88***	-84.47***	-115.8**	-30.58
	(-5.89)	(-4.10)	(-5.32)	(-2.33)
All Gap	-1.199***	-1.050***	-1.328***	-1.122***
	(-4.55)	(-4.05)	(-4.35)	(-3.38)
All Gini	-0.232	0.019***	0.010	0.011
	(-1.65)	(3.45)	(1.31)	(1.54)
Bet Gap	-0.484**	-0.439**	-0.243	-0.238
	(-3.26)	(-3.07)	(-1.46)	(-1.33)
Bet Gini	0.0134	0.0115	0.0165	0.010
	(1.68)	(1.50)	(1.59)	(0.98)

For each measurements, the first row is the coefficient for the variable in interest, second row is the t test statistics. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

1.3.4 Another look at the data

Next, we look at the raw data in an effort to identify changes from pre- to post-merger that led to our estimation results. We first identify the airports that experience entry/exit by the merging carrier(s). Entry is defined as merging carrier(s) operating at the airport for at least 5 months post-merger but none pre-merger. Similarly, exit is defined as merging carrier(s) operating at the airport for least 5 months pre-merger but none post-merger. We find that the merging carrier(s) entered into 6 airports (4 of them in Hawaii), and exited from 2 airports. They also changed their presence significantly (being present for at least 10 more or 10 fewer months post-merger) at 5 airports: 1 of them experiences increase and while 4 experience decreases. For the remaining airports (80 of them), merging carriers "consistently" operated at these airports pre- and post-merger. Next, we focus on these airports.

For each of these airports, we calculate how many routes the merging carriers fly from that airport each month. We then average the number of routes for all pre-merger and post-merger months respectively. Out of the 80 airports, 56 experience reduction in number of routes, for a total reduction of 79.295. 11 airports experience increases, for a total of 4.936.[24] We then consider the airports that both US and HP operated at pre-merger (there are 35 such airports). 32 of them see lower number of routes post-merger, for a total reduction of 59.290. Only 2 airports see more routes, for a total of 1.018. We then consider how the non-merging carriers behaved at these 35 airports where both US and HP operated at pre-merger. Clearly merging carriers reduce their number of routes post-merger, but is this due to merger or is it just time trend? To help answer this question,

[24]13 airports have equal number of routes pre- and post-merger.

we consider the non-merging carriers. For each of the non-merging carriers, we calculate its number of routes, and then average them over the pre- and post-merger months respectively. Afterwards, we add them up across the non-merging carriers. The number of routes by non-merging carriers went down at 4 airports, for a total of 69.489. In contrast, 31 airports experienced increase, for a total of 581.798. Were these increases a response to merging carriers' reducing presence or were the non-merging carriers simply expanding? To get a better idea, we look at airports where merging carriers do not operate on. There are 235 such airports. 67 of them experienced reduction in number of routes by the non-merging carriers, for a total of 60.811, while 142 airports saw more routes post-merger, for a total of 508.087. This seems to suggest that while the merging carriers were contracting in terms of number of routes, non-merging carriers were expanding, wether the merging carriers operated at the airport or not.

Above we have used airport as the unit of analysis and compared number of routes offered at each airport. Next, we use route as the unit of analysis and analyze number of flights per route and arrival delay. We consider only the routes that the merging carriers operated at both pre- and post-merger. For the routes served by both US and HP pre-merger, the number of flights per route went down from 79 to 65. For routes served only by US pre-merger, the comparison is 134 vs. 113. On the other hand, the routes served only by HP pre-merger experienced an increase - number of flights per route went up from 101 to 108. Moving on to non-merging carriers on these same routes, we find that number of flights per route went down from 137 to 114. If we consider only the routes that USHP did not operate on, the change was from 100 pre-merger to 89 post-merger.

We also check arrival delay and find that average arrival delay went down from 11 to 9 minutes on routes where both US and HP operated on pre-merger.[25] On the same routes, arrival delay for non-merging carriers went up from 10.4 to 12.8 minutes.[26]

1.4 Conclusion

We analyze the impact of merger on various measures of product differentiation (vertical and horizontal), using data from the U.S. airline industry. Based on the matching diff-in-diff results, we find that the U.S. Airways and American West merger reduces arrival delay about 0.7 minutes (which benefits consumers). After merger, the merging carrier reduces the number of routes and destinations which it offers non-stop service, and reduces the number of flights per route (all 3 reductions likely will hurt travelers). For horizontal differentiation, we focus on the distribution of scheduled departure times (and thus gaps between neighboring flights) across carriers. We find that merger reduces *Gap_All*, suggesting that flight gaps (in minutes) are more evenly distributed after merger, if we use all flights to calculate the flight gaps. On the other hand, merger seems to have little impact on the other three measures of horizontal differentiation: Gap_Between, Gini_All and Gini_Between. We also conduct several falsification tests and find that our matching method, together with the covariates used for match, works well for some measures of product differentiation but not for others.

Our paper contributes to the literature on the airline industry and the literature on merger impact. We extend the commonly studied on-time performance to product differentiation in general, and identify the impact of merger on various

[25] Arrival delay actually went up from 10 to 12 minutes on routes only US operated pre-merger. On routes where only HP operated on pre-merger, arrival delay went down from 9.7 to 8.2.

[26] If we only consider routes which US and HP do not operate on, arrival delay for non-merging carriers also went up, from 10.6 to 13.5 minutes.

measures of vertical and horizontal product differentiation. Properly evaluating the impact of merger is very important to both academics and policy makers. We propose a matching method which can be used to construct a properly matched group. This matched group then can be compared with the treated group to identify the impact of merger. Our method can be adapted to study other dimensions of merger impact, for example, how merger affect prices. It can also be applied to settings where a market experiences a change other than merger, whether this change is endogenous (driven by the market players themselves) or exogenous (say by regulators).

Chapter 2

Is the Tarmac Delay Rule a Panacea for Passenger Right Protection?

2.1 Introduction

Delay has been a common occurrence in commercial air traffic. Delay can be due to various factors such as bad weather, mechanical problems, or simply because demand for facilities (airport or crew) exceeds their capacity. In the last case, aircrafts may need to wait on the tarmac for their turn to take off. This wait can be quite long during peak hours at busy airports, and can be a lot longer when weather is bad. One may think it is easy to solve this problem by just having the aircraft return to the terminal. However, there are several reasons why this is undesirable. First, there may not be a gate immediately available for the aircraft to come back to. More importantly, turning around forfeits a plane's position in the queue. As a result, the airline industry has seen many flights with extended wait on the tarmac.[1] This has raised public outcries and as a result, U.S. Department of Transportation (hereafter DOT) imposed a new rule to limit airline tarmac delays. In particular, the "new rule prohibits U.S. airlines operating domestic flights from permitting an aircraft to remain on the tarmac for more than three hours without deplaning passengers . . . (DOT 199-09)". The tarmac delay rule initially covers large and medium airports only (not small or non-hub airports), with exceptions for safety and security as determined by the pilot in command, or for significant disruption of airport operations as

[1]See, for example, "Passengers trapped on runway for 8 hours," for a JetBlue flight, www.cnn.com, February 15, 2007.

determined by air traffic control. Otherwise violations of the new rule would lead to hefty financial penalty of up to *$27,500 per passenger*. The intended objective of the rule is to eliminate lengthy tarmac delay and protect passenger rights. The law has largely achieved this goal. In 2009, there were a total of 868 flights with tarmac delay more than three hours. However, this number went down dramatically to 124 flights since the implementation of the tarmac delay rule. The number went down further over time: there were 50 and 42 flights grounded more than three hours in year 2011 and 2012 respectively.

As with every policy, there can be unintended consequences. In this paper, we study the unintended consequences of U.S. DOT's tarmac delay rule. The overall impacts of eliminating lengthy tarmac delay by putting a hard time limit are not as straightforward as it sounds, and concerns towards the possible negative externality arise. It is often argued that tarmac delay is out of an airline's control, for example, in the case of congestion or bad weather. When flights approach the three hour threshold, airlines are more likely to cancel their flights to avoid the fine, and cancellation will lead to even more disruption for passengers who initially are able to maintain their seats and may be able to depart soon. Consumers may have a very different view towards tarmac delay and cancellation, since cancellation involves an extra rearrangement and presumably longer total travel time.

This is not the first study to investigate the tarmac delay rule and flight cancelation. For example, Fukui and Nagata (2014) has confirmed that tarmac delay rule induces "a risk-averse behavior of carriers causes higher flights cancellation rate, and this side effects last at least two years".[2] However, my paper differs

[2] My results also confirm this hypothesis that cancellation rate increases after the implementation of the tarmac delay rule. My results also show that carriers improve the flight rearrangement efficiency once a cancellation occurs.

from earlier studies in several aspects. First, we use more rigorous econometric techniques. The new rule initially only covers large and medium hub airports, so I adopt a difference-in-difference strategy which identifies the difference of covered and uncovered airport for the baseline estimation. Second, we take into account network effects, in particular, by distinguishing between direct cancelation and total cancelation. For a flight with multiple segments, a rule-induced cancelation at a covered airport may lead to subsequent cancelations at rule-uncovered airports (i.e., small or non-hub airports).[3] If one ignores this network effect and uses diff-in-diff method directly by comparing the treatment group (large and medium airports) and control group (small and non-hub airports), the results can be quite misleading. Our results show that for every direct cancelation the tarmac delay rule induces, it leads to two indirect cancelation. Third, we investigate the possibility that the tarmac delay rule may have differential impacts on different carriers, an issue raised by regional carriers. My results did not recover significant difference between the rule's impacts on legacy vs. regional carriers. Fourth, I look at airlines' response to the tarmac delay rule, in particular, their rearrangement efficiency – the ability to reshuffle aircrafts across different routes in its network once cancelation occurs. My results show that the likelihood of subsequent flight cancellations decreases after the new rule which indicates carriers take actions to improve the rearrangement efficiency.

On August 2011, the tarmac delay rule was revised to incorporate small and non-hub airports (DOT, 2011a). As a result all airports were covered by the revision. Given that most of the airlines adopt hub-and-spoke system, and a

[3]Relevant to this policy, DOT divides airports into 4 categories: large, medium, small and non-hub, in decreasing size. A large hub/ median airport refer to an airport that accounts for at least 1.00/ 0.25 percent of the total passenger enplanement in the United States. For example, Oklahoma City and Tulsa airports are small airports while Jackson Hole (Wyoming) and Evansville (Indiana) airports are non-hub airports.

great amount of routes is between large/ median and small/ non-hub airports, this policy extension actually profoundly changes the application of the tarmac delay rule. This paper does not estimate the impact of policy extension because the estimation strategy relies on the valid control and treatment groups. Since the policy extension having all the airports covered by the new rule, it is unlikely to find a reliable control group.

The rest of this paper is organized as follows. Section 2.2 offers more details about the industry and policy background. Section 3 explains the data source and variable construction. I discuss the model and estimation strategy in Section 4. I first quantify the impacts of the tarmac delay rule, and the possible differential impacts on legacy vs. regional carriers. I then analyzes airlines' response in terms of rearrangement efficiency. Section 5 presents the results and explanations. A series of falsification tests are conducted in Section 2.6 and the results suggest that our econometric models have adequately controlled confounding factors. I conclude in Section 2.7.

2.2 Background

Economists have been mostly focused on pricing power in the airline industry. More recently, researchers start to shift to non-price competition such as delays. For passengers who care about total travel time, defined as the *scheduled* departure time to *actual* arrival time, tarmac delay is one of the delay measures they should pay attention to. Airlines are likely to take tarmac time into account when scheduling flights. For instance, flights departing from large and busy airport will be assigned a longer outbound tarmac time than flights departing from smaller, less busy airports. This discrepancy indicates that the airlines will adjust the tarmac time according to the specific characteristics of the routes or airports.

Also, tarmac time can involve significant seasonality, and airlines will schedule longer tarmac time during bad weather seasons such as thunderstorm or snow seasons. With all the adjustments, tarmac delay is still difficult to predict or avoid. Lengthy tarmac delay incidents reached its peak in the summer of 2009. According to the Bureau of Transportation Statistics, more than 500 flights experienced more than three hours of the lengthy tarmac delay that summer (from June to August). DOT issued the tarmac delay rule on December 21, 2009, to be effective on April 29, 2010.

Based on U.S. DOT's documentation, the new tarmac delay rule prohibits U.S. airlines operating domestic flights from permitting an aircraft to remain on the tarmac for more than three hours without deplaning passengers, with exceptions allowed only for safety or security or if air traffic control advises the pilot in command that returning to the terminal would disrupt airport operations. A year later, the tarmac delay rule was expanded to cover not only domestic flights, but also set a four-hour limit to international flights. In this paper, I only estimate the policy impact on domestic flights due to data availability.

The tarmac delay rule also requires the U.S. carriers to adopt contingency plans for lengthy tarmac delays that include provisions for adequate food and water within 2 hours and deplaning of passengers within 3 hours. Once the three-hour threshold is reached, DOT requires the carrier to report the lengthy tarmac delay immediately. Moreover, DOT has a specific record retention requirement that will allow it to look back for a two-year period and determine both the cause of the delay and whether the carrier adequately met its passengers' needs during the incident. If DOT identifies the violation of the rule, carrier may face up to $27,500 per passenger civil penalty. Most of the time, the exact amount of penalty is subject to negotiation. The table below lists the bills issued by DOT up until March 2015.

Table 2.1: DOT bills list

Carrier	Date of Order	Date of Incident	Location	Penalty ($)
American Eagle	11-14-11	05-29-11	ORD	900,000
Jet Blue Airways	08-20-12	03-03-12	JFK	90,000
Pakistan International Airlines	09-19-12	10-29-11	IAD	150,000
Copa Airlines	12-31-12	06-22-12	JFK	150,000
Virgin America	12-31-12	07-18-12	ORD	55,000
United Airlines	02-11-13	05-07-12	ORD	130,000
Caribbean Airlines	03-29-13	08-15-12	JFK	100,000
Air China	05-02-13	07-15-12	JFK	90,000
American Eagle	07-02-13	12-25-12	DFW	200,000
Avianca	08-09-13	08-24-13	MIA	100,000
United Airlines	10-25-13	07-13-12	ORD	1,100,000
Alaska Airlines	11-22-13	05-22-13	PHL	30,000
Qantas Airlines	01-15-14	03-21-13	DFW	90,000
British Airways	04-08-14	11-07-12	EWR	225,000
Air Europa	05-14-14	11-07-12	JFK	140,000
Rouge	10-28-14	01-11-14	BUF	90,000
Southwest	01-15-15	01-02-14	MDW	1,600,000

There have been concerns about the new rule, focusing on its unintended consequences. Many believe that lengthy tarmac delays are caused by a combination of bad weather and the air traffic control (ATC) condition, and these situations in general are out of airlines' control. As a consequence, reduction of lengthy tarmac delays is often associated with increasing flight cancellation. Since the new rule does not impose any restriction on departure delay, it is possible that airlines may leave aircrafts at the gates longer (which increases departure delay) and spend less time on the runway.[4]

The other concern is about the rule's impacts on regional carriers' who typically operate under code-share agreement with legacy other carriers and are usually not responsible for scheduling. To investigate this concern, I also test whether the new tarmac delay lead to differential impact on legacy and regional carriers.

[4] We test this in Section 5 and find there is a significant .08 minutes increase in average delay increase and .59 minutes decrease in taxi out.

2.3 Data

The main data is comprised of On-time Performance (hereafter OTP) and T-100 data from Bureau of Transportation Statistics. I use sample period of 2008 - 2011Q3 to study the impacts of the initial tarmac delay rule before its extension.

OTP data are reported by domestic carriers that account for at least one percent of domestic scheduled passenger revenue. It contains detailed information on a daily non-stop flight's scheduled departure time, scheduled arrival time, departure and arrival delay in minutes, total elapsed time, whether a specific flight is canceled and identification code for each flight. I supplement the on-time performance data with T-100 data, which contains number of seats and passengers at the carrier-route-quarter level. Due to potential punch errors or miscalculation of the tarmac time, I remove the observations with negative "air time" (flight time in minutes). I define a route as a directional airport-pair since the tarmac delay and on-time performance in general can be directional specific even for the same airport-pair.[5] In other words, we treat ATL-ORD and ORD-ATL as different routes because their tarmac times can be quite different. The summary statistics is presented in Table 3.1.

2.3.1 The importance of addressing network effects

The OTP dataset is a disaggregated flight level data, with a unique identifier "tail_num" provided for each aircraft. Now consider a specific aircraft which flies from airport A to B first and then onward to airport C. Suppose that A is a large airport for which the initial tarmac delay rule applies, but B is a small airport

[5] In airline research, we see both directional and non-direction routes, and airport-pair vs. city-pair. For our underlying topic of tarmac delay, it is natural to use directional airport-pairs as routes.

Table 2.2: Summary Statistics for Tarmac Delay Rule Study

	Entire Sample		Pre-policy		Post-policy	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Direct cancellation (ijt)	12.43	43.16	11.504	40.895	13.886	46.462
Total cancellation (ijt)	20.685	80.468	18.838	76.49	23.587	86.276
Cancellation ratio (ijt)	0.395	0.206	0.394	0.209	0.397	0.2
Ave delay increase (ijt)	0.717	1.407	0.681	1.363	0.774	1.472
Taxi out (jt)	14.702	4.307	15.03	4.521	14.181	3.888
HHI (jt)	.0736	.1312	.0715	.1299	.0768	.1331
Market share (ijt)	0.229	0.279	0.218	0.274	0.247	0.287
Total flight(10000) (jt)	1.3557	1.8531	1.3932	1.8786	1.2967	1.8110
N	13999		8554		5445	

not covered. It is probably not surprising that if this aircraft's flight from A to B is canceled to comply with the tarmac delay rule, then its scheduled flight from B to C will be canceled as well. In this sense, tarmac delay rule also affects flights at airports not directly covered by the rule. I call this the network effects. As a result, comparing the trip level total number of cancellation at covered and non-covered airports misleadingly discounts the network effect-induced cancellation and results in underestimating the true policy impact. To properly evaluate the impacts of tarmac delay rule, I need to take into account the cancelations at uncovered airports which are the consequences of rule-complying cancelation in preceding segments.

Aircraft ID	**Total Trip**	**# of cancellation**	**First Cancel Airport**	**Cancel Ratio**
1	4	1	ORD	.25
2	5	1	JFK	.20
3	3	2	IAD	.67
4	6	2	JFK	.33

The "cancellation ratio" is defined as the ratio of total cancellations a given aircraft experience over all the trips it serves per day. For instance, the first row shows within a particular day aircraft 1 services 4 different non-stop directional trips with its first cancellation at ORD, and this aircraft experiences 1 cancellation in total, therefore, the cancellation ratio is .25 (1 total cancellation / 4 total trips). It is possible that due to the network effect the total cancellation is more than 1 trip, which is the case for aircraft 3 and 4. For instance, the cancellation ratio for aircraft 3 is .67 (2 total cancellations / 3 total trips).

To take into account the network effects, I investigate all the routes a given aircraft services per day. The direct cancellation is defined as the first cancellation for an aircraft within a day, and total cancellation is the total number of cancellation an aircraft experiences within a day. Therefore, I'm able to identify all the downstream cancellations associated with the first cancellation from the same aircraft. The "first cancellation" at large or medium airport is assumed to be affected by the policy while the "first cancellation" at small or non-hub airport is not. I aggregate flight level observations into aircraft level data, and assign a binary variable with value 1 if aircraft's first cancellation occur at large or medium airport.

For an aircraft that serves multiple trips, the first cancellation is likely to cause cancelation of subsequent trips. Only the increase in the number of first cancellation is the direct impact of the new rule, while the downstream cancellations should be viewed as the indirect impact. For instance, an aircraft serves a series of trips: A→B→C→D→E, suppose the airline cancels trip B→C to avoid the penalty from the tarmac delay rule, and this is likely to cause the cancellation of downstream trip C→D. I treat the first cancellation B→C as the direct impact of the new tarmac delay rule, while the downstream C→D cancellation is the result of the network effect.

2.3.2 Measuring the Delay Increase

Departure delay is defined as the difference between the scheduled departure time and the actual departure time from the origin airport gate. Since the new rule does not impose any restriction on departure delay, it is possible that airlines may trade tarmac delay with departure delay. That is, carriers may purposefully hold the aircraft at the gate which increase the departure delay to avoid lining aircraft in the queue on the runway when airport is highly congested. When looking at departure delay, the same issue of network effects applies here, i.e., the standard departure delay (as reported in OTP data) is also subject to network effects.

Therefore, in this paper I develop a measure that is independent between trips and is free from the "snowball effect": the "average delay increase" is defined as the different between current trip's departure delay and the preceding trip's arrival delay. For instance, for an aircraft serving trips A→B→C→D→E, if the aircraft's arrival delay at airport B is 10 minutes and next trip's departure delay (which is at airport C) is 30 minutes, then "delay increase" is 20 minutes at airport C for this particular aircraft. While if B's arrival delay is also 10 minutes and C's departure is 10 minutes, then there is no "average delay increase" at airport C.[6]

Therefore, I calculate each flight's "average delay increase" at the origin airport and aggregate into carrier-origin airport-quarter cell to study whether the airlines are more cautious and willing to sacrifice departure delay for the tarmac delay. Although passengers care more about the total travel time or arrival delay they experience, the delay increase is an important measurement from supply side's perspective since it decides whether carriers effectively use the aircraft.

2.4 Estimation strategy

2.4.1 The impact of policy on cancellation

The baseline estimation uses data from 2008 to August 23, 2011. This study focuses on estimating the policy effect on cancellation, differential impact between legacy and regional carriers, as well as carriers' aircraft rearrangement efficiency change.

I implement the difference-in-difference strategy to estimate the short run

[6]If the actual gap is less than scheduled gap, it will be a negative number

policy effect on cancellation before the August 23, 2011 extension:

$$\begin{aligned} Y_{ijt} = \alpha + \mu_t + \gamma_{jt} + \beta_1 Policy_t + \beta_2 Covered_j + \beta_3 Covered_j * Policy_t \\ + \beta_4 X_{ijt} + \epsilon_{ijt} \end{aligned} \tag{2.1}$$

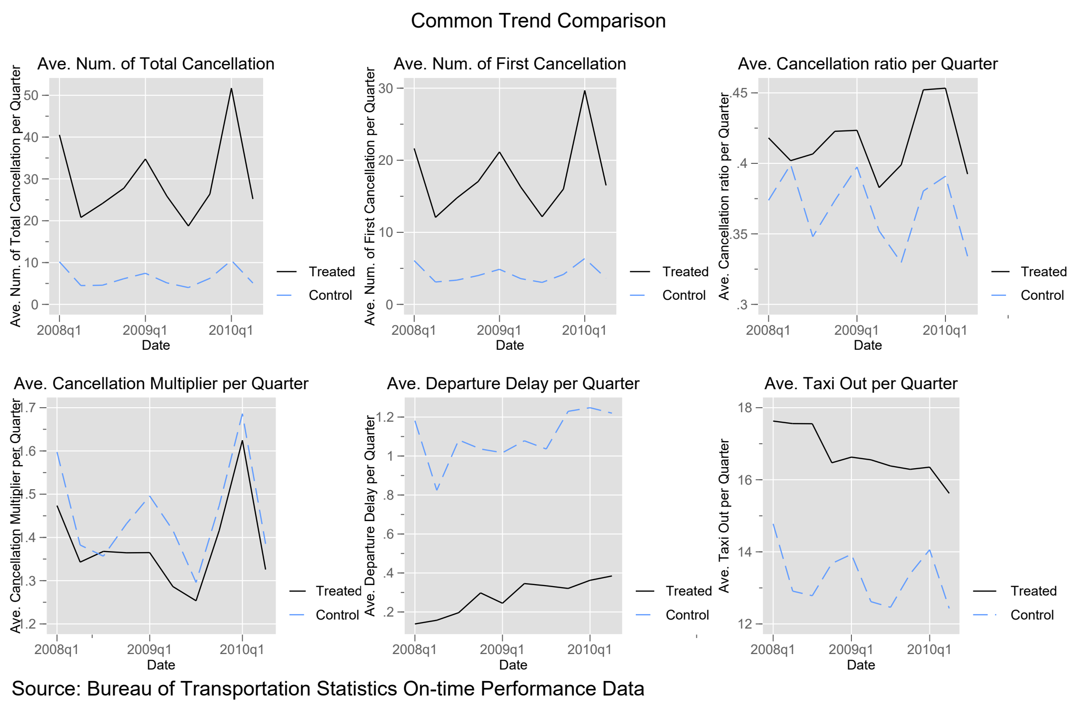
where i denotes the carrier, j refers to the origin airport and t is time (quarter). Y_{ijt} includes number of cancellation without network effect (direct cancelation), with network effects (total cancellation) or delay increase. $Policy_t$ is a binary variable with value 1 to denote the period after policy implementation, which is April 2010. $Covered_j$ is a binary variable to denote whether the origin airport is covered by the initial policy (i.e., large or medium airports), and $Covered_j * Policy_t$ is the interaction term. X_{ijt} is a vector of carrier-origin airport level control variables. I include airport-level, number of passengers based Herfindahl index (HHI) and market share to control for the market structure at the origin airport, I also use total number of flights departing from the airport to control for the traffic volume. μ_t is the year fixed effect and γ_{jt} is airport-quarter fixed effect. Finally, in all models, the standard error is clustered at carrier-airport level to account for the heterogenous correlation over time for each carrier-airport pair.

2.4.2 The validity of diff-in-diff method

One of the key assumptions in the difference-in-difference model is the exogeneity of the policy. In our setting, while the tarmac delay policy targets on tarmac delays longer than 3 hours, it does not impose any restriction on cancellation and average tarmac delay time. The impact on cancellation and average tarmac delay time, which is the focus of this paper, comes from the side-effect of the new rule.

Therefore, the new rule is more likely a "shock" to the cancellation and average tarmac delay time. In addition, the policy is applied at the airport level. As long as an airport is a large or median hub airport, it is automatically covered by the tarmac delay rule so airports can not self select into treated vs. control (or comparison) group. At the airline level, as an important aspect of the consumer protection program, the new policy is mandatory and applies to all airlines. Although exemption requests have been made by several airlines over time, all the exemption requests were rejected after an investigation.[7] Carriers do not have the ability to self-select into treated vs. control group either. Correspondingly, the attrition or selection issue will not bias my estimation results.

Figure 2.1: Common Trend Comparison for Hub and Non-hub Airports



As is common with diff-in-diff analysis, it is important to have similar pre-

[7]Due to a runway closure at JFK airport, JetBlue and Delta Air Lines asked for a temporary exemption from the tarmac delay rule. American Airlines at the same time was also considering applying for the same exemption.

intervention time trend for the treated and control groups (common trend test). To verify the validity of common time trend assumption in DiD, I restrict the sample to pre-policy period and adopt the following estimation:

$$Y_{ijt} = \alpha + \mu_t + \gamma_{jt} + \beta_1 T_t + \beta_2 Covered_j + \beta_3 Covered_j * T_t + \beta_4 X_{ijt} + \epsilon_{ijt}, \tag{2.2}$$

where T_t is a (linear) time trend variable and *Covered* * T_t is the interaction term. For dependent variables (Y_{ijt}), I use direct cancellation (without considering network effects), total cancellation (with network effect), delay increase and cancellation ratio. The control variables (X_{ijt}) are the same as in equation (1). The interaction term ($Covered_j * T_t$) allows me to assess whether the pre-policy time trend is different for covered airports vs. their non-covered counterparts. The results in Table 2.3 indicate that there is not a statistically significant difference between the pre-policy time trends for covered airport-carrier and non-covered airport-carrier pairs for all four dependent variables.

Differential impact on legacy vs. regional carriers

Regional carriers complain that they are typically not in charge of scheduling flights and would have limited ability to reshuffle their flights in the presence of lengthy tarmac delays. As a result, they argue that the policy hurts them more and they should be exempted from the new tarmac delay rule (Answers to Frequently Asked Questions Concerning the Enforcement of the Final Rule on Enhancing Airline Passenger Protections, April 28, 2010). To analyze the validity of this claim, I adopt a diff-in-diff-in-diff (DDD) estimation strategy to investigate whether the policy have differential impact between legacy and regional carriers.

Table 2.3: Difference-in-difference Common Trend Test for Tarmac Delay Rule Study

	(1)	(2)	(3)	(4)	(5)	(6)
	Dir. cancel	**Tot. cancel**	**Ave. delay inc.**	**Taxi out**	**Cancel ratio**	**Cancel Multiplier**
Covered airport * Time trend	0.302	0.0607	.000475	-0.0661*	0.00524**	0.00634
	(0.47)	(0.21)	(.00878)	(-2.48)	(2.80)	(1.00)
Time trend	0.450**	0.228***	.00748	-0.0140	-0.00151	0.00646
	(3.19)	(3.48)	(.00837)	(-0.73)	(-1.23)	(1.50)
Observations	8554	8554	9367	9367	8554	8554

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In particular, I estimate the following model,

$$\begin{aligned} Num\ of\ cancellation_{ijt} = \alpha + \mu_t + \gamma_{jt} + \beta_1 Legacy_i * Covered_j * Policy_t \\ + \beta_2 Covered_j * Legacy_i + \beta_3 Covered_j * Policy_t \\ + \beta_4 Policy_t * Legacy_i + \beta_5 Policy_t + \beta_6 Legacy_i \\ + \beta_7 Covered_j + \beta_8 X_{ijt} + \epsilon_{ijt} \end{aligned} \tag{2.3}$$

The coefficient of interest is β_1, the coefficient for the triple interaction term, which measures the policy's differential impact on legacy vs. regional carriers at large or medium airport after the implementation of policy. $Legacy_i$ is a dummy variable which takes value 1 if the operating carrier i is a legacy carrier (such as American Airline, United Airline, Delta Airline, Hawaii Airline or Alaska Airline etc.). The results are qualitatively the same whether I treat Southwest as a legacy carrier or not.

2.4.3 Rearrangement efficiency

It seems reasonable as to airlines' claim that most of the lengthy tarmac delays are due to either weather or highly congested runway situation which are out of the carriers' control. In the meantime, it is possible that airlines can shuffle aircrafts across routes to minimize downstream cancelations (rearrangement efficiency), once cancellation occurs. In this section, we check whether/how airlines respond to the new policy by improving the flight rearrangement efficiency once cancellation occurs.

Total number of cancellations can be broken down into policy-complying direct cancellation and network effects-induced indirect cancellation. Carriers can improve their flight rearrangement to decrease the likelihood of downstream cancellation so as to minimize the disturbance. I use the aircraft level data to analyze the impact of the new tarmac delay rule on cancellation ratio. The econometric model is as follows,

$$
\begin{aligned}
Cancellation\_ratio_{ijt} = \alpha + \mu_t + \gamma_{jt} + \beta_1 Policy_t \\
+ \beta_2 Covered_j + \beta_3 Covered_j * Policy_t + \beta_4 X_{ijt} + \epsilon_{ijt}
\end{aligned}
\tag{2.4}
$$

The coefficient β_3 tells the mean change of cancellation ratio after implementation of the new tarmac delay rule. I also perform and verify that the common trend test is passed. The decrease in the cancellation ratio means the likelihood of cancelling downstream flights is lower, in other words, an improvement in the flight rearrangement efficiency.

It is possible that rearrangement efficiency (or the change therein) may be associated with the total number of trips an aircraft serves. In particular, it may be easier to reshuffle trips when there are multiple trips left than there are only one or two left. To test this hypothesis, I further divide the data by total number of trips the aircraft serves and then compare the β_3 across different groups.

2.5 Results

2.5.1 Baseline analysis

The new tarmac delay rule increases the cancellation mainly through two mechanism: the first one is the side-effect as the airlines try to protect themselves by cancelling flights to avoid violating the 3 hours threshold. The second channel is the network effects as upstream flight cancellation induces more downstream cancellation. The results of the new tarmac delay rule on total number of cancellation is presented in Table 2.4.

Compared to diff-in-diff results, the OLS result significantly underestimate the impact on direct cancellation, total cancellation and delay increase. The OLS result shows that the new tarmac rule increases roughly 1 direct cancellation per carrier-origin airport-quarter, while the diff-in-diff result shows an increase of 3 flights. Moreover, once the network effects are taken into account the total cancellation significantly increase from 1.057 to 2.861 flights based on OLS result, and rise from about 3.112 to about 6.2 flights based on diff-in-diff results.

Although the intended goal of tarmac delay rule is to protect passenger rights, its indirect negative impact is not negligible. Comparing the results with vs. without network effects also quantifies the "ripple effect" which provides meaningful insight to the regulators: every initial cancellation leads to on average 2 downstream cancellation.

Table 2.4 columns 7-9 show that the impact is not limited to the cancellation. Consistent with the hypothesis, the diff-in-diff results indicate a statistically significant 0.0822 minutes average delay increase due to the new tarmac delay rule at large or medium airports per quarter. The new rule effectively reduces the likelihood of lengthy tarmac delay, but it also increases the delay minutes and arguably increases the total travel time. The average delay increase indicates every time flight passes through a large or medium hub airport it takes longer time to serve a single segment than during the pre-policy period. Although the average increase by 0.0822 seems not a stunningly large number, but based on the summary statistics there are over 13000 flights departure from the airport each quarter on average, the total delay increase is about 1114 minutes. Since the average delay increase is based on each segment and does not account for the "network effects" of delay, from the social welfare point point of view the consumers experience longer than 1114 minutes of delay increase. The last three columns of Table 2.4 reveal the new tarmac delay rule also significantly decrease the average taxi out minutes. The minutes that aircraft spend on the runways is as important as departure delay for the sake of total travel time, I observe about 4% decrease in the taxi out after the policy change.

2.5.2 Differential impacts on legacy vs. regional carriers

In order to better understand whether the new tarmac delay rule impose differential impacts on legacy and regional carriers, I adopt the DDD estimation strategy. The triple interaction term of legacy carrier, large or medium size airport and policy indicator in column 3, 6 and 9 of Table 2.4 demonstrate that the new rule does not impact legacy and regional carriers differentially. The insignificant triple interaction coefficient in column 3 suggests that the new tarmac delay rule imposes similar direct cancellation effect on both legacy and regional carriers. The coefficient in column 6 shows that legacy and regional carriers also experience similar network effects once cancellation occurs. The result also shows both types of carriers experience similar delay increase.

Although legacy carriers typically have more resources and more complicated route networks to reshuffle their flights when needed, they do not seem to have much of an edge over the regional carriers in reducing flight cancellation. Correspondingly, our results suggest that there is no need to over-emphasize the disadvantage of regional carriers in terms of the side-effect of the new rule.

Table 2.4: Estimation Results of Tarmac Delay Rule Study

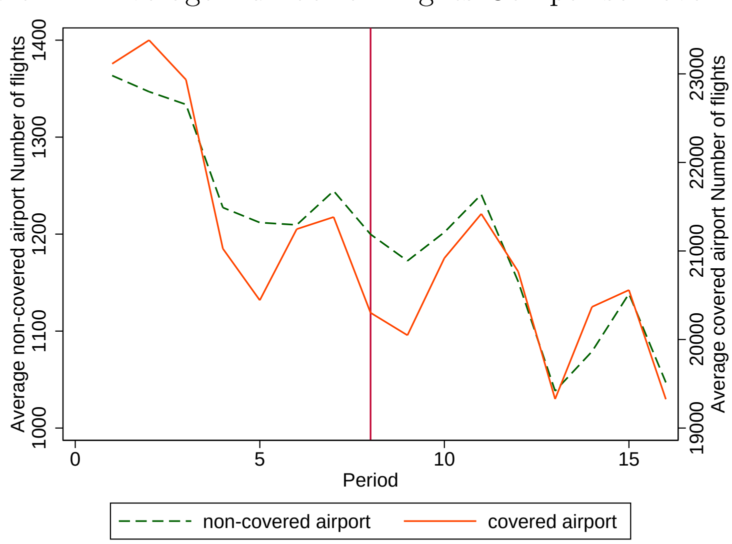
	Direct cancellation			**Total cancellation**			**Ave delay increase**			**Taxi out**		
	(1) OLS	(2) DiD	(3) DDD	(4) OLS	(5) DiD	(6) DDD	(7) OLS	(8) DiD	(9) DDD	(10) OLS	(11) DiD	(12) DDD
Legacy			-1.003			-2.532*			0.234***			1.597***
			(-1.61)			(-2.15)			(3.58)			(5.59)
Policy	1.057*	-0.653	-0.469	2.861**	-0.544	-0.195	-0.0222	-0.0664	-0.0542	0.0392	0.359***	0.250**
	(2.11)	(-1.07)	(-0.75)	(2.95)	(-0.46)	(-0.16)	(-0.71)	(-1.59)	(-1.19)	(0.59)	(4.28)	(2.86)
Covered * Policy		3.112***	2.910*		6.200**	6.436*		0.0822*	0.0841		-0.595***	-0.568***
		(3.38)	(2.35)		(3.27)	(2.52)		(2.10)	(1.95)		(-5.82)	(-4.59)
Legacy * Policy			-0.923*			-1.957*			-0.101			0.483**
			(-2.23)			(-2.56)			(-0.97)			(2.60)
Legacy * Covered			3.116			-0.776			-0.106			-0.754*
			(1.01)			(-0.14)			(-1.46)			(-2.19)
Legacy * Covered * Policy			0.905			0.239			0.0577			-0.211
			(0.60)			(0.08)			(0.54)			(-0.90)
Total flight	-1.548	-0.582	-0.539	1.260	3.184	3.088	-0.161**	-0.135**	-0.128**	1.609***	1.421***	1.458***
	(-0.32)	(-0.12)	(-0.11)	(0.11)	(0.28)	(0.28)	(-3.09)	(-2.73)	(-2.61)	(5.73)	(5.17)	(5.23)
HHI	2.887	2.630	3.105	-4.696	-5.207	-5.456	0.447*	0.442*	0.438*	-1.291**	-1.257*	-1.300**
	(0.45)	(0.41)	(0.47)	(-0.37)	(-0.41)	(-0.43)	(2.39)	(2.36)	(2.38)	(-2.64)	(-2.58)	(-2.89)
Constant	5.865	4.435	4.135	4.928	2.079	3.104	0.183*	0.146	0.0941	13.81***	14.08***	13.72***
	(0.85)	(0.64)	(0.59)	(0.32)	(0.13)	(0.20)	(1.99)	(1.65)	(1.06)	(34.28)	(35.63)	(34.01)
Observations	13999	13999	13999	13999	13999	13999	15265	15265	15265	15265	15265	15265

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.5.3 Rearrangement efficiency

As noted earlier, the new tarmac delay rule led to increase in direct cancellation, total cancellation and the impacts do not seem to differ between legacy and regional carriers. The baseline results suggest that direct cancellation on average will lead to two more cancellations due to network effects. That is, two thirds of the total cancellation is due to network effects. In this section I will investigate whether carriers take actions to reduce the "cancellation network effect" after the policy. I use cancellation ratio and cancellation multiplier to test the change of "cancellation network effect", where cancellation ratio is defined as the ratio of canceled trips over all the trips an aircraft served per day, and cancellation multiplier measures the number of downstream canceled flights over number of first canceled flights.

Figure 2.2: Average Number of Flights Comparison over Time



An important concept as explained in section 3.1 is the "first cancellation", the carriers would repetitively use same aircraft to serve different routes every day to aircraft utilization. The first time an aircraft experiences cancellation would lead to subsequent routes' cancellation, and only the "first cancellation" occurring at large or medium airport is counted as affected by the new tarmac rule.

Table 2.5 reports the results based on OLS, diff-in-diff and triple difference estimation for cancellation ratio and cancellation multiplier. It indicates that after the new tarmac delay rule airlines reduce the cancellation multiplier by 0.0667, but the new rule does not significantly affect the cancellation ratio. Although the new rule arguably increase the indirect cancellation, the decrease of cancellation multiplier lowers the possibility of subsequent cancelations. The triple difference results are insignificant for both cancellation ratio and cancellation multiplier, which indicates the legacy and non-legacy carriers respond similar to the new rule in terms of rearrangement efficiency.

I also check for the case where the initial flight is canceled while the carrier use different aircraft to supplement for the initial canceled flight. If so, then passengers will not experience this cancellation and in my sample I will see two flights from same carrier departing within a short time slot. If a flight is operated by the same carrier as the former canceled flight and the scheduled departure time gap is less than 30 minutes (includes the case that two flights depart at the same time), then the following flight is defined as re-shuffled flight which is used to cover up the initial cancellation. To test this possibility, I calculate the total number of re-shuffled flights and the results are presented in Figure 2.3. We can that only about 1% of flights shows "re-shuffle" behavior if the initial flight is canceled, which suggests that this re-shuffling behavior should not significantly affect the baseline result.

Table 2.5: Cancellation Ratio and Cancellation Multiplier

	Cancel ratio			**Cancel multiplier**		
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	DiD	DDD	OLS	DiD	DDD
Legacy			0.0159			-0.144***
			(1.27)			(-3.65)
Policy	-0.00638	-0.00212	-0.00660	0.0759**	0.112***	0.121***
	(-0.87)	(-0.26)	(-0.77)	(2.85)	(3.47)	(3.57)
Covered * Policy		-0.00775	-0.00496		-0.0667*	-0.0543
		(-1.02)	(-0.57)		(-2.52)	(-1.79)
Legacy * Policy			0.0290			-0.0409
			(1.70)			(-0.78)
Legacy * Covered			0.0594***			-0.0514
			(3.68)			(-1.13)
Legacy * Covered * Policy			-0.0264			-0.0196
			(-1.32)			(-0.34)
Total flight	-0.311**	-0.308**	-0.354***	0.690	0.712*	0.889**
	(-2.86)	(-2.84)	(-3.78)	(1.93)	(1.99)	(2.81)
HHI	-0.0584*	-0.0580*	-0.0520*	0.315	0.318	0.307
	(-2.48)	(-2.47)	(-2.23)	(1.74)	(1.75)	(1.82)
Constant	0.417***	0.420***	0.402***	1.338***	1.368***	1.427***
	(21.23)	(21.16)	(20.10)	(17.34)	(17.80)	(18.50)
Observations	13999	13999	13999	13999	13999	13999

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.6 Falsification Test

One may think that our finding of significant policy impacts on cancelation is not the true policy impact, but rather because I fail to properly control for factors other than the policy (e.g., time trend). This is a legitimate concern for which I perform several types of falsification tests.

The first two falsification tests are similar to each other. In both tests, I restrict the sample to be non-covered airports only (but both before and after policy change). And since they are not covered before or after the policy change, we expect to see no "policy" impacts.

In the first falsification test, I assign small airports to the fake "treated group" and non-hub airports to the control group. The same diff-in-diff strategy is applied to estimate the impacts of this fake "policy change" and the results are

Figure 2.3: Flights with Re-shuffle Behavior

presented in Table 2.6. If our estimation has properly controlled confounding factors then we should see no impacts because there is no real policy change. The results are consistent with this. In column 1, 4, 7 and 10, we do not find significant difference between the fake treated group and the fake control group. In the second falsification test, I sort all the non-covered airports alphabetically. I then assign the first (second) half of airports to fake treated (control) group. The same diff-in-diff strategy is applied with similar results. Results in columns 2, 5, 8 ,11 show there is no significant difference between the fake treated and control group.

Next, I consider both covered and non-covered airports but restrict the sample periods to pre-treatment periods only. I then divide the pre-treatment periods into two period in terms of time, the first half being the "pre-treatment" period, and the second half as the *fake* "post-treatment" period. Since the whole period in this test is before the actual policy change, we would expect to find no policy impact either. Our results are consistent with this. In particular, the interaction

Table 2.6: Falsification Test for Tarmac Delay Rule Study

	Dir. Cancel			**Tot. Cancel**			**Ave Delay Inc.**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Fake	Random	Time	Fake	Random	Time	Fake	Random	Time
Covered * Policy	0.248	0.0865	-1.803	0.533	-0.0127	-2.029	.1199	0.0265	0.0286
	(1.00)	(0.37)	(-1.03)	(1.16)	(-0.03)	(-0.53)	(0.09)	(0.08)	(0.05)
Policy	0.424	0.533*	2.698***	1.440**	1.770***	5.360**	-0.1302	-0.0422	0.2371
	(1.64)	(2.34)	(3.47)	(3.22)	(4.40)	(3.12)	(.09)	(.08)	(1.05)
Observations	6664	6664	8554	6664	6664	8554	7403	7403	9367
	Taxi Out			**Cancel Ratio**			**Cancel Multiplier**		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Fake	Random	Time	Fake	Random	Time	Fake	Random	Time
Covered * Policy	0.0285	-0.118	-0.240	0.00268	-0.0101	0.0192	0.0348	-0.00852	-0.000751
	(-0.50)	(-0.85)	(-1.52)	(0.24)	(-0.92)	(1.74)	(0.72)	(-0.19)	(-0.02)
Policy	0.200	0.214	-0.239	0.0146	0.0212	-0.00179	0.110*	0.135**	0.0730*
	(1.48)	(1.89)	(-1.78)	(1.16)	(1.75)	(-0.21)	(2.24)	(2.59)	(2.45)
Observations	7403	7403	9367	6664	6664	8554	6664	6664	8554

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

term between covered airport and policy interaction term in column 3, 6, 9 and 12 show no significant impacts. These three falsification tests indicates that our econometric models have satisfactorily controlled confounding factors and the policy impacts we find are likely to be the true policy impacts.

2.7 Conclusion, Policy Implication and Discussion

This paper studies the impacts of the U.S. DOT's tarmac delay rule on flight cancellation, departure delay, and average departure tarmac time. This rule aims at reducing the lengthy tarmac delay and protect passengers' rights. However, the policy also has several unintended consequences which end up hurting passengers. First, carriers cancel more flights to avoid the hefty penalty when tarmac time approaches three hours. Second, carriers sacrifice departure delay to avoid having flights on the tarmac when the airport is highly congested.

One key aspect in cancellation and departure delay estimation is the network effect. Every flight is part of the airline's network. Current trip's on-time performance (cancellation in particular) affects subsequent trips. This paper disentangles tarmac delay rule's direct and indirect impacts on cancellation. I also use average delay increase to analyze the effect of new rule on departure on-time performance. The average delay increase is calculated as the difference between current trip's departure delay and preceding trip's arrival delay, which is a trip-independent measure. The diff-in-diff results suggest that the new rule led to a significant 3.112 increase of direct cancellation per (covered) airport-quarter. In addition, I also observe a 0.0822 significant average delay increase per (covered) airport-quarter.

Carriers likely realize the cost of increasing cancellation and take actions to minimize total cancellation (and its corresponding economic impacts). My results show that carriers are able to reduce the cancellation multiplier by 0.0667 which decreases the likelihood of subsequent flight cancellation. As a result, the cancellation network effect only lead to a total of 6.2 flight cancellation per airport-quarter (relative to a 3.1 direct flight cancellation). The positive impacts of the new rule is not limited to the dramatic reduction of lengthy tarmac delay, it also significantly reduces the average departure tarmac time. This paper also addresses the concerns from the non-legacy carriers and shows that the policy has no significant differential impact across legacy vs. regional carriers on direct cancellation, total cancellation, average delay increase and average departure tarmac time minutes. I also show that legacy and non-legacy carriers have similar cancellation ratio and cancellation multiplier, which helps rule out the case that the similar direct and total cancellation increase is because two parties respond to the new policy differently. This paper is not only a retrospective policy impact study, it also shows the necessity of taking cancellation network effects and other side effects into account when making policy decisions.

Chapter 3

Product Differentiation and Quality Uncertainty in the U.S. Airline Industry

3.1 Introduction

Vertical product differentiation occurs when consumers strictly prefer a high-quality product over a low-quality product. On the other hand, horizontal differentiation reflects how consumers have different tastes towards product characteristics. The literature assumes that consumers have full information when making a purchase decision. However, the perfect information assumption is questionable, especially in the airline industry.

On-time performance is one of the most important measures of vertical differentiation. Due to weather conditions, airline traffic congestion, or security threats, only around 70% of flights (based on the years 2004- 2008) are reported as on time. This paper empirically estimates the impact of on-time performance on flight temporal differentiation.

In this paper, the flight temporal differentiation measure is based on the flight departure time. Bester (1998) pointed out two theoretical effects that can influence horizontal differentiation: price competition effect and market share effect. The price competition effect is compromised when consumers are imperfectly informed about product quality. If so, the market share effect will dominate the price competition effect. In order to gain a larger market share, companies locate close to each other and "agglomeration" is the result of market share effect. Therefore, Bester concludes that if quality uncertainty exists then the market will end up with “minimum level of horizontal differentiation”.

Bester’s theoretical conclusion, however, may not be applicable to the airline industry. For instance, the airline industry is not a typical monopoly market, as assumed in both the Hotelling (1929) model and the Bester (1998) model. Borenstein and Netz (1992) pointed out that the real world deviates from most of the model’s assumptions: in the airline industry demand is elastic and passengers are nonuniformly distributed in terms of preferred departure time. In addition, every route is part of a network: scheduling is not entirely determined by the characteristics of a single route. In addition, airlines compete on ticket price, departure time, on-time performance and other factors at the same time.

Since each flight on the route competes with its closest preceding and closest subsequent flights, and all other flights, I construct two measures for horizontal differentiation: time to noon (TTN) and Gini coefficient. Although both measures are based on departure time, TTN emphasises specifically how a given flight competes with all other flights on the route. Gini coefficient, on the other hand, focuses on the differentiation between a flight and its closest neighboring flights.[1] After correcting for potential endogeneity, I find that arrival delay increases flight temporal differentiation on the route, whereas arrival delay decreases the differentiation between a flight and its closest neighboring flights.

In section 2, I review the theories of horizontal differentiation, then discuss the market share effect, price effect and delay cost as the deterministic driving forces of horizontal differentiation in the airline industry. In addition, I review empirical studies of horizontal differentiation and the difficulty of applying differentiation theories to the airline industry. Section 3 covers the data and temporal differentiation measures in the airline industry. Based on the theoretical indication, it is

[1]In this paper, “closest neighboring flights” refers the closest preceding flight and closest subsequent flight.

important to distinguish the three types of flights when discussing the impact of on-time performance on temporal differentiation: own flight, one-sided flight and two-sided flight. Section 4 discusses the estimation strategies. The major empirical concern is the endogeneity issue. Therefore, I use route distance and arrival delay of the previous trip as instrumental variables for temporal differentiation. Results and interpretations are presented in section 5.

3.2 Theories and empirical studies of horizontal differentiation

3.2.1 Theories of horizontal differentiation and quality uncertainty

Hotelling's location theory has been extended and applied to situations where firms have more than one outlet. Following the Hotelling model, Martinez-Giralt and Neven (1988) assume a duopoly market where each firm has two plants, in which firms first choose locations and then compete for price. A key assumption in their model is the quadratic transportation cost assumption. They find that the firms intend to obtain minimum differentiation between their own plants while locating plants away from their competitors. Gabszewicz and Thisse (1986) further relax the quadratic assumption and allow two firms to locate as many outlets as they want to investigate whether Martinez-Giralt and Neven's conclusion is driven by the assumption that only two outlets are allowed. Gabszewicz and Thisse show that the maximum differentiation equilibrium requires competitive pairs of plants to locate evenly over the market space. Their result confirms the maximum differentiation conclusion does not dependent on how many plants each firm has.

Similar to that of d'Aspremont, Gabszewicz, and Thisse (1979), Bester (1998) employed the basic Hotelling duopoly model to show quality uncertainty may lead to agglomeration. In the model, firms choose a location to maximize profit. If a firm moves closer to its competitors, the market share of the moving firm will increase. The opposing effect is the price effect: price competition is intensified once companies are closer to each other. However, the price effect is compromised in the imperfect information framework. Bester (1998) showed that consumers learn about a seller's quality after the purchase and then decide about repeat purchases from the same seller. Unobservable quality reduces the incentives for differentiation by relaxing price competition. In this repeat purchase setting, if the frequency of the repeat purchase is small, then the market share effect is likely to dominate the price effect. As a result, we would observe "minimum differentiation" in the market. In the airline industry, when consumers purchase the ticket they are not able to recognize whether their flights will be delayed. Consumers are able to observe the quality of the on-time performance to some extent, but severe weather conditions, congested airports, and late aircraft issues, make the on-time performance hard to predict.

In this paper, I focus on three different driving forces that potentially affect horizontal differentiation. The first two are universal among the literature. In the price effect, firms intend to maximize the extent of horizontal differentiation to reduce the price competition. However, the market share effect leads to the minimum differentiation. The last driving force is the delay cost. Arrival delay imposes an extra cost on the consumer in addition to the price and adjustment cost, which occurs when actual departure time is different than consumer's scheduled departure time. Consumers need to cover the cost associated with rescheduling their travel plans such as extra car rental fees and hotel re-booking costs. Therefore, the delay cost forces carriers to avoid clustering flights, which works in the opposite direction as the market share effect.

3.2.2 Empirical studies of horizontal differentiation in airline industry

In addition to the large amount of theoretical work on horizontal differentiation, only a few empirical studies investigated temporal differentiation in the airline industry. The most closely related empirical work is from Borenstein (1999). He found a negative relationship between competition and product differentiation after controlling for other factors that might affect departure-time crowding. Borenstein also shows that price is a major factor in deciding the optimal amount of horizontal differentiation. When prices are fixed, there is a strong tendency for firms to crowd departure time. In addition, market structure has a significant impact on horizontal differentiation: departure times are less differentiated if the route is served by competing airlines than if it is served by a single firm.

In the classical Hotelling model, demand is assumed to be perfectly inelastic. Many theoretical studies relax this assumption and extend the model to allow for elastic demand. For instance, Eaton (1972) assumes a linear elastic demand function. He finds that minimum differentiation is reachable with the market length less than a threshold value. Eaton also derives the critical parameters to satisfy the minimum differentiation. However, Eaton believes the equilibrium is not stable since there is no pure Nash equilibrium. In the airline industry, the inelastic demand assumption rarely holds. Instead, the elasticity is likely to vary across market structures. For instance, the route with a dominant proportion of business travelers has relatively inelastic demand, whereas routes with more tourists should be more elastic.

The scheduling of airline flight departures could be analyzed under the horizontal differentiation framework. Analogous to the classical Hotelling model in which the consumers are distributed over a finite line, passengers are distributed over the 24-hour clock in the airline industry. However, passengers are not uniformly distributed.

The airline industry is much more complicated than a duopoly market. However, the flight temporal differentiation shares some common characteristics with the classical Hotelling model in the sense that carriers both compete on ticket price and market share at the same time. Carriers first decide the departure time then compete for the price to maximize profit. Therefore, the two-stage departure time and price competition strategy can be applied to the airline industry.

3.3 Data and measures of temporal differentiation

3.3.1 Data

The estimation period includes 2004-2008. The sample starts from 2004 to avoid the influence of the September 11, 2001 and ends in 2008 to exclude the impact from the airline merger wave. From 2004 to 2008, there is only one large carrier merger: US Airways and America West. This merger has a relatively small impact on the market structure since there are only a few overlapping routes between US Airways and America West. The routes that either US Airways or America West operate on should not experience a drastic change in the market share. In this paper, I use a dummy variable to control for this merger.

The data includes three sources: Bureau of Transportation Statistics on-time performance data, Airline Origin and Destination Survey (DB1B) and Air Carrier

Statistics T-100 domestic segment data. The DB1B data is a 10% sample of airline tickets from reporting carriers, including origin, destination and other itinerary details of passengers information. I further limit the sample to ticket prices higher than 15 U.S. dollars to exclude the tickets purchase by frequent flyer miles. The on-time performance data contains detailed information on a daily non-stop flight's scheduled departure time, scheduled arrival time, both departure and arrival delay minutes, and a unique identification code for an aircraft. Due to the typing errors or miscalculations of the tarmac time, there are some unrealistically lengthy tarmac delays in the on-time performance data. In order to clean the data, I remove the observations with negative air fly times and tarmac delays longer than 10 hours.

I supplement the on-time performance data with T-100 data, which contains quarterly data on the number of seats and passengers. To avoid the contamination from temporary or diverted routes, I restrict the sample to routes on which at least 9 flights operated per quarter. In order to make the estimation manageable, I aggregate data into carrier-route-quarter cells. I estimate least-squares and instrumental variable with fixed effect models, weighting each observation by the number of flights in that cell. Doing so yields the same estimation results as using disaggregated flight-level data. I follow Borenstein and Rose (1994) to define a route as a non-stop directional combination of origin and destination pair since the arrival delay and related horizontal differentiation is route-level phenomena. For instance, the Atlanta (ATL)-Chicago O'Hare (ORD) route is treated differently than route Atlanta (ATL)-Chicago O'Hare (ORD).

3.3.2 Measures of temporal differentiation

I follow the classic Hotelling model and assume the consumer utility function as:

$$U(a, p_i, q_i, y_i, k, d_i) = q_i - p_i - t(a_i - y_i)^2 - k(d_i) \tag{3.1}$$

The main assumption is that consumers are *a prior* uninformed about q_i of route i and they can only learn the actual quality after experience the service of the flight trip.[2] Consumers are fully informed with the price (p_i). Consumer's ideal departure time is a_i and actual departure time is y_i, t indicates the degree of temporal differentiation, and $t(a - y_i)^2$ denotes consumer's adjustment cost. k is the consumer's sensitivity to the delay, and d_i is the delay of route i.

In order to analyze the impact of on-time performance on temporal differentiation, I construct two measures for temporal differentiation. Both measures are based on the flight departure time gap. The first measure, TTN, is defined as:

$$Time\ to\ noon = |departure\ time - 12|. \tag{3.2}$$

12 o'clock is assumed to be the middle point of a day. Each flight competes with all the other flights on the route. A TTN value of 0 means minimum differentiation and a value of 12 denotes maximum differentiation. That is, the smaller the TTN, the stronger the incentive for a carrier to compete for higher market share.[3] Bester (1994) believes that the quality uncertainty mitigates horizontal differentiation, which means a large arrival delay is associated with a small TTN. However, some assumptions that Bester (1994) made are not applicable in the

[2] See detailed discussion: Bester (1998), Quality uncertainty mitigates product differentiation.
[3] This means in the equation (1) the $t(a_i - x - i)^2$ is small, and the market share effect dominates the price effect if flights are clustered at 12 o'clock.

airline industry.

The second measure is the Gini coefficient. Consider a given route-day which contains more than 2 flights. For each flight (except the earliest and latest flights of the day in terms of the scheduled departure time), I calculate two gaps: *preceding gap* is the gap between this flight and its closest preceding flight in terms of scheduled departure times; *subsequent gap* is the gap between this flight and its closest subsequent flight.[4] Suppose that a given route-day has N flights that are not the earliest/latest flights of the day.Some of the N flights may have the same scheduled departure time (e.g., American and United may both offer a 10 a.m. flight on a given route). Then each of these flights has either preceding or subsequent gap at zero. Suppose a flight is departing at 9 a.m., which has a closest preceding flight departing at 7 a.m. and has a closest subsequent flight departing at noon. Then there are a 2-hour closest preceding gap and a 3-hour closest subsequent gap for the flight. We then have a total of $2N$ gaps.Sorting both preceding and subsequent gaps in an ascending order, we can then calculate the Gini coefficient.

With all the ranked departure gaps, I follow the usual expression to calculate the Gini coefficient:

$$G = \frac{2}{(2N)^2\bar{x}} \sum_{i=1}^{2N} i(x_i - \bar{x}) \tag{3.3}$$

where G is the Gini coefficient,[5] $\bar{x}$ is the mean value of the ranked departure gap distribution, N is the number of flights in the sample, $2N$ is the number of ranked departure gaps, and x_i is the i^{th} ranked departure gap, where i = 1, ...,

[4]Such gaps are not calculated for the earliest and latest flights of the day since they miss either a preceding or subsequent flight by definition. If multiple flights are tied for the earliest/latest flight of the day, then the gaps are not calculated for any of these tied flights.

[5]The first and last flight in a day are excluded from the estimation since they only have one neighboring flight. Therefore, only one gap can be calculated. I exclude first and last flight in a day when calculate the Gini coefficient

$2N$.[6]

The Gini coefficient is one of the most commonly used measures of inequality, and it has many advantages that are particularly useful to measure flight temporal differentiation. Since it is a scale-independent measure, there is no need to normalize by the number of flights on the route. Therefore, the Gini coefficient can be used to compare the temporal distribution across different markets. In addition, the Gini coefficient is easy to interpret. A large Gini coefficient indicates a greater inequality of the temporal distribution between a flight and its closest neighboring flights.

In sum, TTN measures how a flight differs from all the other flights on the route, whereas the Gini coefficient indicates how a flight differs from its closest preceding and closest subsequent flights.

3.3.3 Own flight v.s. one-sided flight v.s. two-sided flight

It is important to differentiate three types of flights when discussing the flight temporal differentiation. Borenstain and Netz (1999) partition horizontal differentiation into average within- and between- differentiation according to whether each pair of flights are scheduled by the same airline or is scheduled by different airlines. I follow the similar idea and further separate three types of flight: own flight, one-sided flight, and two-sided flight.

I partition flight types based on the characteristics of its closest preceding and closest subsequent flights. Any given flight should fall into one of the three categories: first is "own flight", which is when both closest preceding and closest subsequent flights are scheduled by the same carrier. Since its closest neighboring

[6]See detailed discussion from Karagiannis and Kovacevic (2000), a method to calculate the jackknife variance estimator for the Gini coefficient.

flights are *both* operated by the same carrier, the market share effect should have a small impact on its temporal differentiation. The second one is called "one-sided flight". If a flight's closest preceding or closest subsequent flight is scheduled by a different carrier (not both flights), then it falls under the "one-sided flight" category. Theoretically, the "one-sided flight" engages in minimum differentiation with flights that are operated by competitors. On the other hand, the one-sided flight maximally differentiates from flights under the same carrier. The last type is "two-sided flight". Both the closest preceding and closest subsequent flights are operated by competitors in this case. Studying the impact of the on-time performance on flight temporal differentiation without comparing these three cases is misleading because different types of flights are subjected to different levels of market share effects, price effects, and network effects. For instance, on-time performance compromises price effect and enhances market share effect. However, "own flight" is less affected by market share effect compared with "one-sided flight" and "two-sided flight". Therefore, on-time performance imposes a smaller impact on horizontal differentiation for "own flight" than the others.

3.3.4 Summary statistics

Table 3.1 presents the summary statistics for measures of temporal differentiation, including each of the samples that I study from 2004-2008. TTN is based on minutes.

I provide the comparison of different flight types for both temporal differentiation measures. The preliminary evidence shows that "own flight" group's TTN is significantly smaller than "one-sided flight" and "two-sided flight". This finding shows that when a flight and its closest neighboring flights are operated by the same carrier, the market share effect dominates the price effect.

The Gini coefficient shows the same pattern as TTN across different flight types.[7] The "own flight" group has a significantly smaller Gini coefficient than the other two cases. One explanation is that "one-sided flight" and "two-sided flight" carriers have stronger incentive to schedule flights closer to competitor and increase the chance to sell to more consumers. The "own flight" group, which both the closest preceding and closest subsequent flights are served by the same carrier, has limited incentive to "steal" consumers from the closest neighboring flights. Figure 3.1 presents the histogram of the Gini coefficient for all three types of flights. The "own flight" Gini coefficient is clustered around 0.152 and has a short tail. The "one-sided flight" and "two-sided flight" have a similar shape. This also supports the conjecture that "own flight" does not compete intensively with its closest neighboring flights. Both the summary statistics and the histogram show that there are few extreme values for the Gini coefficient.

In addition to the market share and price effect, the delay cost also has a significant impact on the flight departure schedule. In particular, a large hub airport is much more crowded than small airports.[8] We should expect a stronger market share effect, price effect, and a larger delay cost in large hub airports. However, it is unclear which effect dominates. Based on Table 3.1, both TTN and Gini coefficients are larger in large hub (origin) airports than others.

Arrival delay denotes on-time performance, and is measured by the difference between scheduled arrival time and actual arrival time. Arrival delay is set to zero for any flights which arrive early. I use the total number of flights on a route

[7] To make the calculation manageable, the Gini coefficient is based on randomly selected 5th, 15th and 25th day of the month. After restricting the data, I have roughly 500,000 observations.

[8] Federal Aviation Administration (FAA) defines a large hub if the airport accounts for least 1 percentage of annual passenger boarding. Based on this definition, the following airports are categorized as large hub airport: ATL, ORD, LAX, DFW, DEN, JFK, SFO, LAS, PHX, IAH, CLT, MIA, MCO, EWR, SEA, MSP, DTW, PHL, BOS, LGA, FLL, BWI, IAD, SLC, MDW, DCA, HNL, SAN and TPA.

Table 3.1: Summary statistics

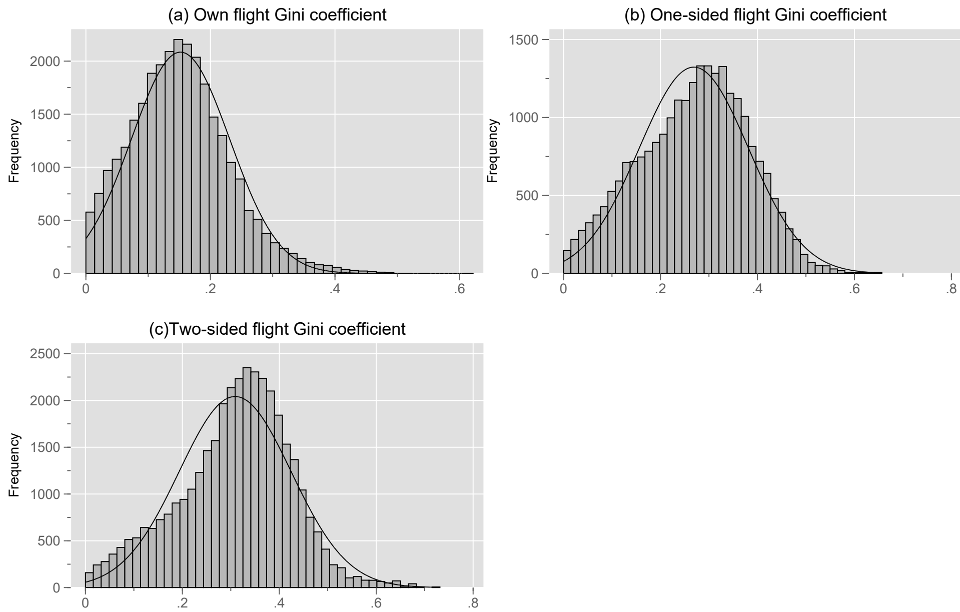
Dependent variable				Control variable			
Variable	**Mean**	**Std. Dev.**	**N**	**Variable**	**Mean**	**Std. Dev.**	**N**
Time to noon (entire sample)	192.495	78.83	67579	Arrival delay (mins)	14.583	9.242	67579
Time to noon (non-large hub)	175.947	72.516	24571	Departure delay (mins)	12.831	8.458	67579
Time to noon (large hub)	201.949	80.711	43008	Ticket price (U.S. Dollars)	203.758	82.600	67579
Time to noon (0-25%)	204.535	83.053	16895	Load factor	0.743	0.114	67579
Time to noon (25%-50%)	195.257	81.331	16895	# of carriers	3.514	1.95	67579
Time to noon (50%-75%)	194.519	84.287	16896	# of flights on route	686.223	371.153	67579
Time to noon (75%-100%)	175.667	61.649	16893	# of flights by carrier on route	357.313	248.205	67579
Gini coefficient (entire sample)	0.27	0.116	67579	Market share (ratio)	0.538	0.335	67579
Gini coefficient (non-large hub)	0.237	0.11	24571	HHI	0.614	0.248	67579
Gini coefficient (large hub)	0.289	0.115	43008	Route distance (miles)	749.934	560.059	67579
Gini coefficient (own flight)	0.152	0.079	29217	Arrival delay of previous trip(mins)	11.531	8.645	58206
Gini coefficient (one-sided flight)	0.269	0.113	24496				
Gini coefficient (two-sided flight)	0.309	0.116	36426				

a. Source: data is from U.S. Bureau of Transportation Statistics.

b. Time to noon is measured by minutes.

c. Arrival delay is measured as the difference between scheduled arrival time and actual arrival time. The early arrival of the flight is recorded as 0 for arrival delay. Load factor is defined as the ratio of number of passengers over total number of seats. # of carriers measures number of unique operating carriers on the route. # of flights on route refers to total number of flights on the route. # of flights by carrier on route measures the number of flights operated by a given carrier. Market share is calculated as the ratio of number of passengers served by a given carrier over the total number of passengers served on the route per quarter. Herfindahl index (HHI) measures the market concentration level, which is calculated by squaring the market share for each carrier and then summing the squares.. Arrival delay of previous trip is defined as the arrival delay of the previous trip which is served by the same aircraft.

Figure 3.1: Histogram of Gini Coefficient by Flight Type



a. Source: data is from U.S. Bureau of Transportation Statistics, the figure is based on author's own calculation.
b. To make the calculation manageable, the Gini coefficient is based on randomly selected 5th, 15th and 25th day of the month.
c. "Own flight" is defined as both its closest preceding and closest subsequent flights are scheduled by the same carrier. "one-sided flight" is defined that a flight's either the closest preceding or closest subsequent flight is scheduled by a different carriers (not both flights). "two-sided flight" is defined as both closest the preceding and closest subsequent flights are operated by competitors.

to account for the traffic volume at each airport. Number of flights by carrier on route controls for the number of flights on the route served by the carrier. I also include market structure variables to control for the route characteristics. Market share is calculated as the ratio of the number of passengers served by the carrier over the total number of passengers. Herfindahl index (HHI) is based on the number of passengers, and is calculated by squaring the market share for each carrier and then summing the squares. I use the load factor, which is calculated as the number of passengers over the total number of available seats, to capture the impact of the capacity constraint on strategic flight departure scheduling. If the flight is nearly full, the incentive to locate the flight close to its competitor's is reduced.

3.4 Estimation strategies

In this subsection, I study the impact of on-time performance on flight temporal differentiation. To gain some insight into the relationship, I use a panel fixed effect model:

$$
\begin{aligned}
Y_{ijt} = {} & \alpha + \mu_t + \gamma_{ij} + \beta_1 delay_{ijt} + \beta_2 fare_{ijt} + \beta_3 load\ factor_{ijt} + \beta_4 \#\ carriers_{jt} \\
& + \beta_5 tot.\ flights(r)_{jt} + \beta_6 tot.\ flights(c-r)_{ijt} + \beta_7 mkt.\ share_{ijt} \\
& + \beta_8 HHI_{ijt} + \epsilon_{ijt}
\end{aligned}
\tag{3.4}
$$

where i denotes the carrier, j refers to the directional route and t denotes time. Y_{ijt} includes 2 different types of temporal differentiation measures: TTN and Gini coefficient.[9] μ_t is the year and quarter fixed effect, which is used to control the time trend change and seasonality, γ_{ij} is carrier-route fixed effect to control for the time-invariant unobservable for each carrier-route pair. Finally, in all models, the standard error is clustered at carrier-route level to account for the correlation in the standard errors over time for each carrier-route pair. The estimation is in the log-log form.[10]

I use the arrival delay to measure the on-time performance as a proxy for quality uncertainty. The endogeneity issue is the major concern of the panel fixed effect model. Although the panel fixed effect model controls the time-invariant unobservable factors, the time-variant unobservable factors will bias the weighted ordinary least square estimation. The second concern is reverse causality: a route

[9]Since the Gini coefficient is bounded between 0 and 1, the estimate results would be biased if the predicted Gini coefficient is less than 0 or greater than 1. The result confirms the predicted value is also bounded between 0 and 1, which indicates the validity of the regression.

[10]It is easier to interpret the result as the elasticity than the level change. The main results are robust to linear and log-linear functional form.

with "agglomerated" departure flights is more likely to experience a delay than a route with evenly distributed departure flights.

To handle the endogeneity issue, I use two instrumental variables for the arrival delay: route distance and arrival delay of the previous trip. Route distance is predetermined for all the flights on the route, and therefore, is uncorrelated with the flight departure time. Aircraft designed for middle-haul routes serve fewer trips per day compare to those designed for short flights. For instance, it takes 1.5 hours for a short-haul aircraft to serve route Washington D.C. (DCA)- Boston (BOS) and it can serve at least 5 similar routes per day. However, it takes 3.5 hours for a middle-haul aircraft to serve route Washington D.C. (DCA)- Dallas (DFW) and it can only serve 2 similar routes per day. One of the major causes of delay is "late aircraft delay".[11] As the route distance becomes longer and the aircraft can only serve fewer trips per day, the likelihood of experiencing "late aircraft delay" is lower.

The second instrument is the arrival delay of the previous trip. Each flight is a part of the airline network and any single aircraft is scheduled to serve multiple trips. The on-time performance of the current trip is significantly affected by the punctuality of the previous trip served by same aircraft. Therefore, I use the arrival delay of the previous trip as an instrumental variable for the current trip's on-time performance.[12]

[11] Based on the Bureau of Transportation Statistics on-time performance data, 76.05% of total flights are reported on time, the major causes of delay are national aviation system delay (7.82%), aircraft arriving late delay (6.98%), air carrier delay (6.04%), weather delay (0.93%) and security delay (0.06%).

[12] I use "tail_num" in the Bureau of Transportation Statistics on-time performance data to pin down a unique aircraft.

3.5 Results

3.5.1 Baseline estimation result

I include both OLS and IV results for TTN and Gini coefficient measures/outcomes. Table 3.2 reports the baseline regression results. In addition to explaining the economic intuition of the instruments in section 4, I also implement the following IV tests: under-identification test, weak identification test, and over-identification.[13] All three tests are passed at least at the 10% significance level for all the regressions. I control for variables that affect TTN, time-invariant unobservable factors, and time trend, the OLS result still underestimates the impact of arrival delay on TTN. After correcting for endogeneity, the estimated coefficient of arrival delay is about twice the OLS estimate. This result indicates that a 1% increase in the arrival delay leads to 14.2% increase in the TTN.

The increase in the arrival delay compromises the price competition and firms have a stronger incentive to compete for market share. This forces carriers to locate their flights close to 12 o'clock in an effort to sell to more consumers.[14] At the meantime, the arrival delay increases the delay cost, which reduces the clustering incentive and drives flights away from 12 o'clock. The estimation result supports the conjecture that the delay cost dominates the market share effect. Column 3 and 4 compare the delay impact of non-large and large hub airports. Since the on-time performance is generally worse in large hub airports, the delay

[13]The under identification test checks whether the instruments are relevant. The null hypothesis is no correlation. The weak identification test checks whether instruments are correlated with the regressors but weakly. The null hypothesis is the instrumental variable and endogenous regressors are weakly correlation. The Hansen J over identification test checks the validity of the over-identifying restrictions. "It tests whether the restrictions implied by the existence of more instruments than endogenous regressors are valid" (P., Parentea and S., Silvac, 2012). The null hypothesis the instruments are coherent with each other.

[14]According to the Hotelling model, if the market share effect dominates price effect firms will locate in the middle of space and experience minimum differentiation.

Table 3.2: Baseline results

	Time to noon				Gini coefficient			
	Whole sample		Non-large hub	Large hub	Whole sample		Non-large hub	Large hub
	(1) OLS	(2) IV	(3) IV	(4) IV	(5) OLS	(6) IV	(7) IV	(8) IV
Arrival delay	0.0749***	0.142***	0.100***	0.169***	0.00768	0.0121*	0.00478	0.0179**
	(21.38)	(27.43)	(13.74)	(24.26)	(1.51)	(1.73)	(0.39)	(2.16)
Ticket price	0.00204	0.000807	-0.0148	0.0133	-0.0516***	-0.0517***	-0.0754**	-0.0332*
	(0.23)	(0.09)	(-1.14)	(1.20)	(-2.95)	(-3.10)	(-2.56)	(-1.69)
Load factor	-0.0744***	-0.108***	-0.101***	-0.117***	-0.0266	-0.0305	-0.0205	-0.0324
	(-5.56)	(-8.40)	(-4.78)	(-7.31)	(-1.25)	(-1.51)	(-0.59)	(-1.30)
# of carriers	0.00259	0.00241	-0.00217	0.00360	0.0568***	0.0566***	0.0811***	0.0457***
	(0.63)	(0.62)	(-0.30)	(0.77)	(7.79)	(8.15)	(6.23)	(5.65)
# of flights on route	0.264***	0.248***	0.321***	0.207***	0.641***	0.642***	0.693***	0.608***
	(20.98)	(20.86)	(15.86)	(14.08)	(23.44)	(24.65)	(15.44)	(19.20)
# of flights by carrier on route	0.00355	0.0129	-0.00385	0.0212	-0.160***	-0.162***	-0.127***	-0.181***
	(0.32)	(1.24)	(-0.23)	(1.62)	(-9.77)	(-10.32)	(-5.97)	(-8.70)
Market share	0.0145***	0.0138***	0.00696	0.0171***	0.0493***	0.0492***	0.0297**	0.0581***
	(3.73)	(3.67)	(1.12)	(3.74)	(3.84)	(4.00)	(2.57)	(3.35)
HHI	-0.0251**	-0.0352***	-0.0427**	-0.0340***	-0.226***	-0.226***	-0.190***	-0.247***
	(-2.24)	(-3.33)	(-2.40)	(-2.58)	(-11.75)	(-12.31)	(-7.34)	(-9.87)
Observations	67381	65823	23786	42037	67412	65852	23799	42053
Under identification test		Y	Y	Y		Y	Y	Y
Weak identification test		Y	Y	Y		Y	Y	Y
Over identification test		Y	Y	Y		Y	Y	Y

a. All models are weighted by the number of flights. All the variables are in the log-log form. The instruments are route distance and arrival delay of previous trip.
b. The under identification test checks whether the instruments are relevant. The null hypothesis is no correlation. “Y” is used to denote the rejection of null hypothesis at 10% significance level.
c. The weak identification test checks whether instruments are correlated with the regressors but weakly. The null hypothesis is the instrumental variable and endogenous regressors are weakly correlation. “Y” is used to denote the rejection of null hypothesis at 10% significance level.
d. The Hansen J over identification test checks the validity of the over-identifying restrictions. The null hypothesis the instruments are coherent with each other. “Y” denotes fail to reject the null hypothesis at 10% significance level.
e. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

cost is accordingly larger. The comparison is consistent with the baseline result that arrival delay increases the temporal differentiation measured by TTN and the impact is stronger at the large hub airport. The results indicate a 1% increase in arrival delay causes the gap between 12 o'clock and flight departure time to increase by 17%.

Column 5 and 6 report the impact of the arrival delay on the Gini coefficient based on the entire sample. The IV-based estimated coefficients of arrival delay are positive and statistically significant, roughly twice as large as the coefficient from OLS estimate. This implies the arrival delay decreases the flight temporal differentiation when measured by Gini coefficient. One might argue that this result is inconsistent with the temporal differentiation measured by TTN: the arrival delay forces carriers to schedule flights away from 12 o'clock and increases temporal differentiation. The Hotelling model assumes a duopoly model and each firm is allowed to locate only one plant. Arrival delay increases the temporal differentiation from the entire route's perspective, which is measured by TTN. But it decreases the differentiation between the flight and its neighboring flights, which is measured by Gini coefficient. In sum, arrival delay drives the flight departure time away from 12 o'clock but it also increases the incentive to compete with its closest neighboring flights (either preceding or subsequent flight).

The estimated impact on Gini coefficient is statistically significant at the large hub airport, but not at the non-large hub airport. I argued that in the large hub airport there are more competitors on the route. Therefore, the carriers have stronger desire to engage in a strategic scheduling. In this way, the product quality uncertainty has a more significant overall impact.

3.5.2 Time to noon across different market structure

As indicated by theoretical models, the level of flight departure differentiation varies across different market structures. In this subsection, I present evidence to show the impact of on-time performance on TTN across different market structures.

The Herfindahl index (HHI) is a commonly used measure for market concentration. According to different quartile of HHI (25%, 50%, 75% and 100%), I separate the whole sample into 4 sub-samples. The market concentration of the route, which dictates whether or not a route is in a given sample, is possibly determined in part by variables that are part of the residual in Equation (4). The right-hand-side variables would be orthogonal to the residuals in a "complete" sample, but it is possible that they are correlated with the residuals in the nonrandom sub-samples. To handle this truncation issue, I first implement the Heckman sample selection equation, then apply the instrumental variable approach to estimate Equation (4). The estimation results are reported in Table 3.3. In all four sub-sample regressions, the impact of on-time performance on TTN is positive. However, the coefficient of arrival delay is decreasing monotonically as the market concentration increases. The results show that, going from the first quartile to the fourth quartile the coefficient decreases from 18.3% to 6.11%. Estimation supports the maximum differentiation hypothesis when the on-time performance worsens, but the impact is larger when the market concentration is close to perfect competitive market.

One of the explanations is that in the perfectly competitive market, consumers are more sensitive to the delay cost since they have more options to choose from. The delay cost dominates the market share effect, which drives carriers schedule flights away from the 12 o'clock. However, as the market concentration increases and carriers have larger market power, the concern of increased delay cost becomes less important for carriers.

3.5.3 Impacts by different flight types

Gini coefficient is based on the departure time between a given flight and its closest neighboring flights (both preceding and subsequent flights). It measures how a flight differentiates from its nearest flights. Whether the flight and its closest neighboring flights are operated by the same carriers determines how they compete with each other. For instance, "own flight" has little incentive to compete intensively with the closest neighboring flights since all the flights are from the same airline. This indicates a small Gini coefficient. This is supported by the summary statistics in Table 3.1. Therefore, it is important to compare three different flight types.

Table 3.4 reports the impact of arrival delay on the Gini coefficient. The Gini coefficient measures the departure inequality of a flight and its closest neighboring flights, so an increase in the Gini coefficient implies a flight and its closest neighboring flights are more unevenly distributed, in another word, less differentiated. In all regressions, the estimated coefficients of arrival delay are positive.

First two columns report the results for "own flight", the weighted OLS and IV regression estimation have similarly sized estimated coefficient. Flights in the "own flight" group have less incentive to compete for market share against the closest neighboring flights. Therefore, the on-time performance does not significantly affect the temporal differentiation of "own flight" type. However, for a "one-sided" type of flight, the carrier has a strong incentive to schedule the flight close to the competitor's and away from its own flight as an effort to

Table 3.3: Time to noon results

Time to noon	(1) 0-25%	(2) 25%-50%	(3) 50%-75%	(4) 75%-100%
Arrival delay	0.183***	0.166***	0.146***	0.0611***
	(15.66)	(14.60)	(15.40)	(9.70)
Ticket price	0.0643***	0.0407*	-0.00529	-0.0332***
	(3.58)	(1.87)	(-0.32)	(-2.67)
Load factor	-0.0756***	-0.0566**	-0.0615**	-0.0708***
	(-3.05)	(-1.96)	(-2.52)	(-3.92)
# of carriers	-0.00403	0.00376	0.0113	-0.00422
	(-0.33)	(0.39)	(1.44)	(-0.91)
# of flights on route	0.176***	0.268***	0.245***	0.339***
	(7.50)	(11.82)	(11.54)	(8.62)
# of flights by carrier on route	0.0216	0.00301	-0.0124	0.00811
	(1.16)	(0.13)	(-0.60)	(0.21)
Market share	0.0299**	0.0314***	0.0151**	0.0422***
	(2.49)	(2.96)	(2.12)	(3.22)
HHI	-0.0697***	0.0797*	-0.0204	0.0339
	(-2.68)	(1.95)	(-0.76)	(0.73)
Observations	16049	15617	15618	16177
Under identification test	Y	Y	Y	Y
Weak identification test	Y	Y	Y	Y
Over identification test	Y	Y	Y	Y

a. All models are weighted by the number of flights. The results are based on IV regression, in a panel fixed effect setting. All the variables are in the log-log form. The instruments are route distance and arrival delay of previous trip.
b. The under identification test checks whether the instruments are relevant. The null hypothesis is no correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.
c. The weak identification test checks whether instruments are correlated with the regressors but weakly. The null hypothesis is the instrumental variable and endogenous regressors are weakly correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.
d. The Hansen J over identification test checks the validity of the over-identifying restrictions. The null hypothesis the instruments are coherent with each other. "Y" denotes fail to reject the null hypothesis at 10% significance level.
e. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: Gini coefficient results

Gini coefficient	Own flight (1) OLS	(2) IV	One-sided flight (3) OLS	(4) IV	Two-sided flight (5) OLS	(6) IV
Arrival delay	0.0164	0.0163	0.0149*	0.0282**	-0.000833	0.00851
	(1.24)	(0.96)	(1.90)	(2.36)	(-0.18)	(1.20)
Ticket price	-0.116**	-0.115***	-0.0510*	-0.0505**	-0.00231	-0.00319
	(-2.45)	(-2.58)	(-1.89)	(-2.03)	(-0.13)	(-0.19)
Load factor	-0.0692	-0.0703	-0.0321	-0.0379	-0.0200	-0.0189
	(-1.40)	(-1.54)	(-0.85)	(-1.09)	(-0.95)	(-0.97)
# of carriers	0.0249	0.0249	0.00350	0.00327	0.0427***	0.0439***
	(1.46)	(1.55)	(0.26)	(0.26)	(4.37)	(4.83)
# of flights on route	-0.785***	-0.787***	0.478***	0.478***	0.343***	0.340***
	(-8.27)	(-8.81)	(11.92)	(12.87)	(13.95)	(14.71)
# of flights by carrier on route	1.030***	1.032***	-0.181***	-0.181***	0.0317***	0.0331***
	(9.46)	(10.08)	(-6.05)	(-6.53)	(2.67)	(2.96)
Market share	0.114***	0.114***	0.00610	0.00564	0.00661	0.00643
	(2.70)	(2.87)	(0.46)	(0.46)	(1.10)	(1.16)
HHI	-0.148*	-0.148*	-0.0319	-0.0329	-0.153***	-0.145***
	(-1.82)	(-1.95)	(-1.08)	(-1.20)	(-7.32)	(-7.81)
Observations	29109	28665	24468	23893	36337	34943
Under identification test		Y		Y		Y
Weak identification test		Y		Y		Y
Over identification test		Y		Y		Y

a. All models are weighted by the number of flights. All the variables are in the log-log form. The instruments are route distance and arrival delay of previous trip.

b. The under identification test checks whether the instruments are relevant. The null hypothesis is no correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.

c. The weak identification test checks whether instruments are correlated with the regressors but weakly. The null hypothesis is the instrumental variable and endogenous regressors are weakly correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.

d. The Hansen J over identification test checks the validity of the over-identifying restrictions. The null hypothesis the instruments are coherent with each other. "Y" denotes fail to reject the null hypothesis at 10% significance level.

e. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

"steal" consumers from the competitor. The estimated coefficients are presented in columns 3 and 4. Both weighted OLS and IV coefficients for the arrival delay are positive and statistically significant. After correcting for the endogeneity, an 1% increase of arrival delay will lead to 2.82% increase in the Gini coefficient. The increased departure gap inequity implies that the arrival delay enhances the market share effect and relaxes the price effect. As a consequence, the market share effect dominates the price effect and delay cost. Based on the coefficient estimated in columns 5 and 6, the on-time performance has no significant impact on Gini coefficient for "two-sided" type flights. When a flight's both proceeding and subsequent flights are from competitors, on-time performance will not affect flight departure differentiation.

3.5.4 Robustness tests

To gain some insight into the robustness of the conclusion, I use the departure delay as the alternative proxy for on-time performance. Table 3.5 reports the impact of departure delay on TTN across different levels of HHI. The result is consistent with using arrival delay: all the estimated coefficients of on-time performance are positive and significant, and the magnitude of the impact is decreasing with the increased market concentration level.

Table 3.6 presents the impact on Gini coefficient of using departure delay, the results for "own flight" and "two-sided flight" are almost the same as the arrival delay based estimation. However, the "on-sided flight" is less robust: the estimated coefficient is statistically insignificant and the magnitude is smaller than arrival delay based measure. This inconsistency implies that consumers and carriers care more about the arrival delay than departure delay.

Table 3.5: Robustness check of time to noon

Time to noon	(1) 0-25%	(2) 25%-50%	(3) 50%-75%	(4) 75%-100%
Departure delay	0.167***	0.151***	0.133***	0.0595***
	(15.68)	(14.72)	(15.49)	(9.71)
Ticket price	0.0559***	0.0415*	-0.00618	-0.0342***
	(3.15)	(1.91)	(-0.38)	(-2.75)
Load factor	-0.0823***	-0.0658**	-0.0617**	-0.0747***
	(-3.32)	(-2.29)	(-2.53)	(-4.12)
# of carriers	-0.00406	0.00483	0.0112	-0.00382
	(-0.33)	(0.50)	(1.47)	(-0.82)
# of flights on route	0.180***	0.274***	0.253***	0.341***
	(7.59)	(12.05)	(12.04)	(8.70)
# of flights by carrier on route	0.0217	0.00614	-0.0143	0.00714
	(1.16)	(0.27)	(-0.70)	(0.19)
Market share	0.0314***	0.0302***	0.0158**	0.0438***
	(2.65)	(2.74)	(2.25)	(3.39)
HHI	-0.0613**	0.0754*	-0.0253	0.0306
	(-2.37)	(1.85)	(-0.94)	(0.66)
Observations	16051	15604	15604	16172
Under identification test	Y	Y	Y	Y
Weak identification test	Y	Y	Y	Y
Over identification test	Y	Y	Y	Y

a. All models are weighted by the number of flights. The results are based on IV regression, in a panel fixed effect setting. All the variables are in the log-log form. The instruments are route distance and arrival delay of previous trip.
b. The under identification test checks whether the instruments are relevant. The null hypothesis is no correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.
c. The weak identification test checks whether instruments are correlated with the regressors but weakly. The null hypothesis is the instrumental variable and endogenous regressors are weakly correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.
d. The Hansen J over identification test checks the validity of the over-identifying restrictions. The null hypothesis the instruments are coherent with each other. "Y" denotes fail to reject the null hypothesis at 10% significance level.
e. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6: Robustness check of Gini coefficient

	Own flight		One-sided flight		Two-sided flight	
	(1)	(2)	(3)	(4)	(5)	(6)
Gini coefficient	OLS	IV	OLS	IV	OLS	IV
Departure delay	0.0148	0.0154	0.00987	0.0101	0.00187	0.00764
	(1.07)	(0.95)	(1.29)	(1.41)	(0.42)	(1.21)
Ticket price	-0.116**	-0.115***	-0.0512*	-0.0513**	-0.00230	-0.00315
	(-2.46)	(-2.59)	(-1.90)	(-2.06)	(-0.13)	(-0.19)
Load factor	-0.0693	-0.0711	-0.0307	-0.0311	-0.0216	-0.0193
	(-1.40)	(-1.56)	(-0.82)	(-0.89)	(-1.02)	(-0.99)
# of carriers	0.0250	0.0251	0.00351	0.00357	0.0425***	0.0439***
	(1.47)	(1.56)	(0.26)	(0.29)	(4.35)	(4.84)
# of flights on route	-0.784***	-0.787***	0.479***	0.481***	0.343***	0.340***
	(-8.26)	(-8.81)	(11.93)	(12.93)	(13.93)	(14.73)
# of flights by carrier on route	1.029***	1.031***	-0.181***	-0.181***	0.0317***	0.0331***
	(9.45)	(10.07)	(-6.04)	(-6.53)	(2.67)	(2.96)
Market share	0.114***	0.115***	0.00631	0.00634	0.00656	0.00649
	(2.70)	(2.88)	(0.47)	(0.51)	(1.09)	(1.17)
HHI	-0.148*	-0.148*	-0.0315	-0.0310	-0.153***	-0.146***
	(-1.82)	(-1.94)	(-1.07)	(-1.14)	(-7.34)	(-7.82)
Observations	29103	28665	24462	23896	36288	34928
Under identification test		Y		Y		Y
Weak identification test		Y		Y		Y
Over identification test		Y		Y		Y

a. All models are weighted by the number of flights. All the variables are in the log-log form. The instruments are route distance and arrival delay of previous trip.
b. The under identification test checks whether the instruments are relevant. The null hypothesis is no correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.
c. The weak identification test checks whether instruments are correlated with the regressors but weakly. The null hypothesis is the instrumental variable and endogenous regressors are weakly correlation. "Y" is used to denote the rejection of null hypothesis at 10% significance level.
d. The Hansen J over identification test checks the validity of the over-identifying restrictions. The null hypothesis the instruments are coherent with each other. "Y" denotes fail to reject the null hypothesis at 10% significance level.
e. All models cluster standard error at carrier-route level. The t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

3.6 Conclusion

Due to the weather, security, mechanical issues and other factors, on-time performance is unpredictable. Meanwhile, the punctuality is one of the most important quality measures in the airline industry. In the airline industry, the delay cost reduces consumer's welfare which is not at all accounted for in Hotelling based theoretical models. Under the quality uncertainty framework, I assume three driving forces that can affect flight temporal differentiation: market share effect, price effect, and delay cost.

In this paper, I use route distance and arrival delay of a previous trip to instrument for the arrival delay. The OLS with fixed effect model is likely to underestimate the impact of on-time performance on flight temporal differentiation. After correcting for the endogeneity, I find that the arrival delay significantly increases the flight departure temporal differentiation from the entire route's perspective, which contradicts Bester's (1994) minimum differentiation conclusion. The impact varies across different market structures: arrival delay has a larger impact when the market concentration is low.

I also investigate how a flight's departure time differs from its closest preceding and closest subsequent flights. For the "one-sided flight" group, on-time performance significantly reduces the differentiation between the flight itself and its closest neighboring flights. This finding indicates the quality uncertainty enhances the market share effect when one of the closest neighboring flights is served by the competitor, leading to lower differentiation.

References

[1] Alexis, D., J. Sekhon. (2013) "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *Review of Economics and Statistics*, 95(1), 932-945.

[2] d'Aspremont, C., Gabszewicz, J.J., Thisse, J.-F., 1979. "On Hotelling's stability in competition." *Econometrica*, 17, 1145–1151.

[3] Austan, G., Chad Syverson. (2008) "How do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." *The Quarterly Journal of Economics*, 123 (4): 1611-1633.

[4] Bester, 1998. "Quality uncertainty mitigates product differentiation." *The RAND Journal of Economics*, 29, 828-844.

[5] Bishop, J., N. Rupp, B.Zheng., 2011. "Flight Delays and Passenger Preferences: An Axiomatic Approach." *NBER Working Paper*

[6] Borenstein, S., Rose, N.L., 1994. "Competition and price dispersion in the US airline industry." *Journal of Political Economy*, 102, 653–683.

[7] Borenstein, S., Netz, 1999. "Why do all the flights leave at 8 am?: Competition and departure-time differentiation in airline markets." *International Journal of Industrial Organization*, 17, 611-640.

[8] Brueckner, J., 2002. "Airport Congestion When Carriers Have Market Power." *The American Economic Review*, 92, 1357-1375.

[9] Brueckner, J., P. Spiller., 1994. "Economies of Traffic Density in the Deregulated Airline Industry." *Journal of Law and Economics*

[10] Brueckner, Ricardo, 2006. "Airline schedule competition: product-quality choice in a duopoly model." *working paper.*

[11] Christopher, M., S. Todd., 2003. "Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil." *American Economic Review*, 93, 1194-1215.

[12] Christou, Vettas, 2005. "Loction choice under quality uncertainty." *Mathematical Social Sciences*, 50, 268–278.

[13] Ciliberto, F., J. Williams., 2010. "Limited Access to Airport Facilities and Market Power in the Airline Industry." *Journal of Law and Economics*, 53, 467-495.

[14] Cochran, W., Rubin, D., 1973. "Controlling bias in observational studies." *Sankyha.*, Vol. 35: 417—446.

[15] Dai, M., Q. Liu, K. Serfes, 2014. "Is the Effect of Competition on Price Dispersion Non-Monotonic? Evidence from the U.S. Airline Industry." *Review of Economics and Statistics*, 96(1), 161-170.

[16] Daniel, J., 1995. "Congestion Pricing and Capacity of Large Hub Airports: A Bottleneck Model with Stochastic Queues." *Econometrica*

[17] Denzau, A., Kats, A., Slutsky, S., 1985. "Multi-agent equilibria with market share and ranking objectives." *Social Choice and Welfare*, 2, 95–117.

[18] Diamond, A., Sekhon J., 2013. "Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies." *The Review Economics and Statistics*, 95(3): 932-945.

[19] Eaton, C.B., Lipsey, R.G., 1976. "The non-uniqueness of equilibrium in the Loschian location model." *American Economic Review*, 66, 77–93.

[20] Eaton, C.B., Lipsey, R.G., 1975. "The principle of minimum differentiation reconsidered: Some new developments in the theory of spatial competition." *Review of Economic Studies*, 42, 27–49.

[21] Federal Aviation Administration, 2014. "The Economic Impact of Civil Aviation on the U.S. Economy."

[22] Forbes, S., 2008. "The effect of air traffic delays on airline prices." *International Journal of Industrial Organization.*

[23] Forbes, S., M.Lederman, 2010. "Does vertical integration affect firm performance? Evidence from the airline industry." *RAND Journal of Economics*, 41, 765-790.

[24] Gerardi, K., A. Shapiro, 2009. "Does competition reduce price dispersion? New evidence from the airline industry." *Journal of Political Economy*, 117, 1-37.

[25] Goldberg, David E., 1989. "Genetic Algorithms in Search, Optimization, and Machine Learning." Reading, MA: Addison-Wesley.

[26] Gravelle, Sivey, 2010. "Imperfect information in a quality-competitive hospital market." *Journal of Health Economics*, 29, 524–535.

[27] Guido W. Imbens, 2009. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review." *The Review of Economics and Statistics*, 86(1): 4–29.

[28] Harumi, I., D. Lee, 2005. "Assessing The Impact of the Septermber 11 Terrorist Attacks on U.S. Airline Demand." *Journal of Economics and Business*

[29] Hassin, O., O. Shy., 2000. "Code-Sharing Agreements, Frequence of Flights, and Profits in the Airline Industry." *Working Paper*

[30] Heckman, J., Ichimura, H., Todd, P., 1997. "Matching as an econometric evaluation estimator: evidence from evaluating a job training programme." *Review of Economic Studies.*, Vol. 64 (4): 605—654.

[31] Heckman, J., Ichimura, H., Todd, P., 1998. "Matching as an econometric evaluation estimator." *Review of Economic Studies.*, Vol. 65 (2): 261—294.

[32] Heckman, J., LaLonde, R., Smith, J., 1999. "The economics and econometrics of active labor market programs." *In: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A. North-Holland, Amsterdam, pp.*, 1865—2097.

[33] Heckman, J.J., 1977. "Sample selection bias as a specification error." *Econometrica*, 47, 153–161.

[34] Hotelling, H., 1929. "Stability in competition." *The Economic Journal*, 39, 41–57.

[35] James J. Heckman, H. Ichimura, P. Todd, 1997. "Matching as an Econometric Evaluation Estimator." *The Review of Economics and Statistics*, 65, 261-294.

[36] Jasjeet S. Sekhon, 2011. "Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R." *Journal of Statistical Software*, 42(7).

[37] Johnson, T., I. Savage., 2006. "Departure delays, the pricing of congestion, and expansion proposals at Chicago O'Hare Airport." *Journal of Air Transport Management*, 12, 182–190.

[38] Martinez-Giralt, X., Neven, D.J., 1988. "Can price competition dominate market segmentation." *Journal of Industrial Economics*, 36, 431–442.

[39] Mazzeo, M., 2003. "Competition and Service Quality in the U.S. Airline Industry." *Review of Industrial Organization*, 22: 275-296.

[40] Mebane, Walter R. Jr., Jasjeet S. Sekhon, 2011. "Genetic Optimization Using Derivatives (GENOUD)." *Journal of Statistical Software*, 42(11).

[41] Paulo, J.M.C. Santos, 2012. "A cautionary note on tests of overidentifying restrictions." *Economics Letters*, 115, 314–317.

[42] Prince, J., D. Simon. “The Impact of Mergers on Quality Provision: Evidence from the Airline Industry.” forthcoming, *Journal of Industrial Economics*.

[43] Rich N., J. Sheffield, 2009. “Matching with Time-Series Cross-Sectional Data.” *working paper*.

[44] Rosenbaum, Paul R., Donald B. Rubin, 1983. “The Central Role of the Propensity Score in Observational Studies for Causal Effects.” *Biometrika*, 70(1): 41–55.

[45] Rosenbaum, Paul R., Donald B. Rubin, 1985. “Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score.” *American Statistician*, 39(1): 33-38.

[46] Donald B. Rubin, 1979. “Using Multivariate Sampling and Regression Adjustment to Control Bias in Observational Studies.” *The American Statistical Association*, 74: 318-328.

[47] Donald B. Rubin, 1980. “Bias Reduction Using Mahalanobis-Metric Matching.” *Biometrics*, 36(2): 293-298.

[48] Donald B. Rubin, 1976b. “Multivariate Matching Methods That are Equal Percent Bias Reducing, II: Maximums on Bias Reduction for Fixed Sample Sizes.” *Biometrics*, 32 (1): 121–13.

[49] Donald B. Rubin, 2001. “Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation.” *Health Services and Outcomes Research Methodology*, 2(1): 169-188.

[50] Rupp, N., D. Owens, L. Plumly, 2006. “Does Competition Influence Airline On-time Performance?” *Advances in Airline Economics: Competition Policy and Antitrust*, Vol. 1.

[51] S. Forbes, M. Lederman, 2010. “Does vertical integration affect firm performance? Evidence from the airline industry.” *RAND Journal of Economics*, 41(4): 765-790.

[52] Smith, Jeffrey, P. Todd, 2005a. “Does Matching Overcome LaLonde’s Critique of Nonexperimental Estimators?” *Journal of Econometrics*, 125 (1–2): 305–353.

[53] Snider, C., W. Williams., 2015. ”Barriers to Entry in the Airline Industry: A Multi-Dimensional Regression-Discontinuity Analysis of AIR-21” *The Review of Economics and Statistics*, vol. 97-5, 1002-1022.

[54] Stavins, J., 2001. "Price discrimination in the airline market: the effect of market concentration," *Review of Economics and Statistics*, 83, 200–202.

[55] T. Beatty, J. Shimshack, 2001. "School buses, diesel emissions, and respiratory health." *Journal of Health Economics*, 30(5), 987–999.

[56] Zhong Zhao, 2004. "Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence." *Review of Economics and Statistics*, 86(1), 91-107.

Appendix

3.7 Instruction on how the data is cleaned and organized

The on-time performance data provides information for each flight. In order to make the estimation manageable, we typically aggregate the flights into carrier-route-month cell by weighting each observation by the number of flights in that cell. Recall that we use directional routes, so $A \to B$ is a different route than $B \to A$. Since after the merge all the flights are operated under US Airways' name, we create a carrier code "USHP" to denote both US Airways (US) and America West (HP) pre-merger and the merged carrier post-merger. US Airways is concentrated on the East Coast while America West is on the West Coast. These two carriers have few overlapping routes. Assigning a common carrier code will have negligible impact on the estimation results. After using a common carrier code for both US Airways and America West and treating them as one carrier throughout all the time periods, we are able to make pre- and post-merger comparisons. Next, we describe how we derive our measures of vertical and horizontal differentiation.

Vertical differentiation measures

Our first measure of vertical differentiation is arrival delay. This is observed for each flight. We calculate its average at the carrier-route-month level. Our second measure is number of flights per-route, which can be obtained by directly counting the number of flights at the carrier-route-month level. We also derive the number of destinations, at the carrier-route-month level. For each month, we count how many destinations a carrier has flights originating from each of the two end airports of the routes. We then calculate the geometric mean and use it as the number of destinations for that carrier-route-month. Number of routes is calculated at the carrier-month level. For each carrier-month, we count the number of distinct routes which the carrier operates flight on.[15]

Horizontal differentiation measures

We construct several measures to capture the distribution of flight scheduled departure times. For each flight on a given route-day, we first calculate the gap (in minutes) between it and the closest flight on the same route-day. We consider two variations depending on whether there is restriction on the carrier identity of the closet flight.

Gap/Gini between flights using all flights

In this setting, we impose no constraint on the carrier identify of the closest flight. Therefore, we have a gap measure for each flight on a given route-day, call it gap_all_{mjt}, where m is flight, j is route and t is day. We then square gap_all_{mjt} for each flight and calculate the average across flights on the same route-day, call it $SqGap\_All_{jt}$. $Gap\_All_{jt'}$, our dependent variable, is the simple average of $SqGap\_All_{jt}$ across days of the month. We also use $Gini$ coefficient to measure flight distribution. Instead of calculate Gini coefficient at route-day and then average them at the monthly level, [16] In order to calculate meaningful Gini index, we restrict sample to routes with at least 3 flights per day.

Gap/Gini between flights from different carriers

For this measure, we restrict the gap to be between this flight and a flight of a different carrier. This first rules all monopoly routes, being served by a single

[15] Even though number of routes is at carrier-month level, in the estimation we run it as if it were at the carrier-route-month level, i.e., repetition across routes for the same carrier-month. This allows us to take advantage of the variations explanatory variables across routes and allows us to use route fixed effects as well.

[16] This is mostly to ease the computational burden. If we include all gap_all_{mjt} in a month, and directly calculate $Gini\_All_{jt'}$ (at the route-month level) Calculating Gini at the route-day level take significantly longer time.

carrier. For remaining routes, some flights may still be dropped. Each flight has up to two neighboring flights (possibly more if several flights have the same departure time). We only include flights which have at least one neighboring flight from a different carrier. For each of these flights, we calculate the shortest gap between this flight and a neighboring flight from a different carrier, call it $gap\_between_{mjt}$, where m is flight, j is route and t is day. We then square $gap\_between_{mjt}$ and calculate the average across flights on the same route-day, call it $SqGap\_Between_{jt}$. $Gap\_Between_{jt'}$ (t' for month), our dependent variable, is the simple average of $SqGap\_Between_{jt}$ across days of the month. Similar to $Gini\_All_{mjt'}$, for any route-month, we group all $gap\_Between_{mjt}$'s within that month and on that route, and directly calculate $Gini\_Between_{jt'}$.