ABSTRACT

HE, QINGXIN. Three Essays on Market Structure: An Analysis of Price Discrimination and Marketable Byproducts. (Under the direction of Dr. Xiaoyong Zheng).

In the US-China international flight market, airlines sell the same ticket at two different prices, one at the published fare level and the other with deep discounts. Though the former price is available to all travelers, the latter can only be accessed by consumers who purchase through travel agents. This provides us a rare opportunity to examine how airlines exploit imperfect information and consumers' search costs as a fence to segment consumers and price discriminate. Using a unique dataset collected from this market, we find evidence that third-degree price discrimination with respect to costly consumer search increases with market competition.

While the price dispersion of airline tickets across consumer groups positively relates to market competition, the relationship between price discrimination within a consumer group and market competition unnecessarily follow the same pattern. When airlines cannot segment consumers into smaller groups, they offer a menu of price-ticket quality/restriction combination to make the consumers self-select, and practice second-degree price discrimination. This part of my dissertation offers new evidence on the relationship between competition and second-degree price discrimination in the airline industry, using the same unique micro dataset on the flight market between the US and China. Consistent with the conventional wisdom, the results show that, for most of the ticket restrictions/qualities (e.g., Advance, Duration, Weekend, Stops and Direct), marginal prices and hence second-degree price discrimination decrease with market competition. This indicates that the capability for

firms to price discriminate against consumers within the same group is deteriorated when there are more competitors.

Energy and environmental regulations often result in the generation of unforeseen marketable by-products that can reduce costs of compliance. The second part of my dissertation evaluates a unique case in which the marketability of by-products produces positive external benefits. Specifically, this paper focuses on flue-gas desulfurization (FGD) gypsum by-product produced by coal-fired electric plants using sulfur dioxide (SO2) scrubbing technology to comply with the emission standards set by the 1990 Clean Air Act Amendment (CAAA). Intuitively, the ability to market FGD gypsum provides incentives for electricity generating plants to operate their scrubbers more efficiently by providing gypsum revenues, and lowering gypsum disposal costs. Increased operating efficiency creates external benefits attributed to reduced SO2 emissions that would otherwise not have occurred. In order to estimate the impact of FGD gypsum marketability on SO2 emissions, a unique longitudinal database is constructed that merges electricity plants' boiler-level SO2 emissions data from the U.S. Environmental Protection Agency (EPA) with data on scrubber operations and coal quality available from the U.S. Energy Information Administration (EIA). Boiler fixed-effect estimations suggest annual SO2 emissions decline by 3,494 tons as a result of gypsum marketing. This corresponds with a 39% average reduction in SO2 emissions at boilers marketing FGD gypsum. The total benefits associated with the reduced SO2 emissions are estimated between $147.1 billion and $350.6 billion (2011$), or between 0.9% and 2.1% of the total benefits of the U.S. CAA from 1990 to 2010.

Three Essays on Market Structure: An Analysis of Price Discrimination and Marketable Byproducts

by
Qingxin He

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Economics

Raleigh, North Carolina

2013

APPROVED BY:

Dr. Xiaoyong Zheng Committee Chair	Dr. Stephen E. Margolis
Dr. Robert G. Hammond	Dr. Thayer Morrill

DEDICATION

给我的爸爸妈妈和我爱的人们。

To my parents, and all the people I love.

BIOGRAPHY

Qingxin He was born in Xi'an, Shaanxi Province of China on October, 18, 1984. She started to play violin since seven years old and had obtained the Lv.9 (10 as the highest) violin certificate of the Music Association of China in Junior High. She lived in Tai'an at the base of Mount Tai (the mount of China) with her parents until she left for the U.S. to pursue a PhD in economics after her honorable graduation from the Shandong Economic University. With the help from the alumni of North Carolina State University, and the Chinese Students and Scholars Friendship Association at NCSU, she adapted herself to the new environment in perspective of study and life. During her doctoral study, she had more comprehensive understanding in economics and found her interest in applied microeconomics. Her fields of specialization are industrial organization and environmental economics. While she has diverse interests, most of her current research focuses on the overlap between the two fields.

ACKNOWLEDGMENTS

North Carolina State University is a great university where I spend about six of my prime years over. I study, grow and get to know the world and myself better. Thanks for the great graduate program it provides, I gained comprehensive understanding in the fields of economics I would like to continue my research, I also acquired and strengthened the ability to teach myself. There are so many people have been helping me along the process of pursuing PhD that I cannot fully express my gratitude and appreciation towards them within the limited space. I would like to particularly express my thanks to my dissertation committee comprised of Drs. Xiaoyong Zheng, Stephen Margolis, Robert Hammond and Thayer Morrill.

I would first recognize my advisor Dr. Xiaoyong Zheng, for his earnest and diligent advising on my dissertation. I first expressed my interest in price discrimination in airline market in Dr. Zheng’s empirical industrial organization class, with the encouragement and stepwise guidance provided by Dr. Zheng, I started to explore the uniqueness of the international flight market between the US and China in perspective of market structure, pricing, consumer types and policies. Thanks for the unique price data source provided by Dr. Zheng, I get the chance to examine my research of interest, and fulfill my willingness to be a researcher committed in studying price discrimination and market structure empirically. Dr. Zheng is always responsive in regards to research issues. He spent countless hours in instructing and discussing with me about my dissertation, as well as reviewing my research. I am extremely grateful for his time and effort devoted to my dissertation and our co-authored work.

Dr. Margolis mentors me in such a way that economic intuition is valued. After talking to him in regards to my research, I have a clearer idea about the big picture of my research. His enlightening discussion on the cause and effect in market structure, competition, antitrust and most of the other topics of industrial organization was instrumental in expanding my knowledge of the filed due to its broad coverage of major research areas of industrial organization.

I would like to thank Dr. Hammond for the incredible help in providing suggestions on the latest empirical techniques that relates to my research. He also gives me insightful comments and corrections for my dissertation. The suggestions on phrasing economic rationale intuitively are especially appreciated.

I would also like to thank Dr. Morrill for his help in providing a theoretical economist's perspective on my dissertation research. He also offered great advice on the way I present my paper.

My parents unconditionally love, support and help me all the time. They do everything they can to support my decision and all they hope is that I am happy. I appreciate all of their understanding for the time that I have spent away from home during the important holidays and family get-together, since it is impossible to make transoceanic travel during the academic semester. My mom always gives me detailed tips in taking caring myself in all matters, my dad give me great suggestions in how to make a balance in daily life and my study.

My fiancé Jonathan Lee is also an economist and he is my best friend. Thank him for the suggestions in regards to environmental/energy economics which is my second research of interest. We share similar research of interest in economics and we spend plenty of time in discussing our research and economic issues. He spent countless hours in preparing me for presentation and offered great tips in regards. I am so happy that we have so many things in common both in daily and profession life. I also enjoyed the time we spend on the co-author research.

TABLE OF CONTENTS

LIST OF TABLES ix
LIST OF FIGURES xi

Chapter 1. The Effect of Competition on Price Discrimination in International Flight Market between the U.S. and China 1

1. Introduction 1

2. The International Flight Market between the United States and the People’s Republic of China 5

3. Data 11

3.1 Ticket Data 11

3.2 Official Airline Guide (OAG) Data 13

3.3 Bureau of Transportation Statistics (BTS) Data 16

3.4 Census Data 17

3.5 Summary Statistics 18

4. Empirical Models 19

4.1 Identification...21

5. Results 25

5.1 Baseline Results...26

5.2 Main Results..30

5.3 Robustness Check..32

6. Conclusions 34

REFERENCES..36

Chapter 2. Using Search Costs as Another Fence for Market Segmentation: Evidence from the US-China Flight Market..48

1. Introduction 48

2. The International Flight Market between the US and China 51

3. Data 59

3.1 Ticket Data 59

3.2 Official Airline Guide (OAG) Data 62

3.3 Bureau of Transportation Statistics (BTS) Data 63
3.4 Census Data 65
3.5 Summary Statistics 66
4. Empirical Analysis 68
4.1 Identification 68
4.2 Empirical Models 73
4.3 Endogeneity 75
5. Results 77
5.1 Baseline Results 78
5.2 Main Results 81
5.3 Robustness Check 1 82
5.4 Robustness Check 2 83
6. Conclusions 84
REFERENCES 86
Chapter 3. When a Negative Externality Turns Positive: An Analysis of Flue-gas Desulfurization By-products 108
1. Introduction 108
2. Theory 110
2.1 Wallboard Market 113
2.2 Electric Utility Profit Maximization 115
3. Data 121
4. Results 125
4.1 Net Emissions as a Function of Gypsum Sales 127
4.2 Reduced Emissions due to Operating Efficiency 129
5. Conclusions 134
REFERENCES 137
APPENDICES 151
Appendix A 152

LIST OF TABLES

Table I-1: Summary Statistics..39

Table I-2: Regression Results without Interaction Terms 40

Table I-3: Averages of Estimated Coefficients for Year-Month Dummy Variables...............41

Table I-4: Regression Results with Interaction Terms _OLS...42

Table I-5: Regression Results with Interaction Terms_2SLS...43

Table I-6: Selected Results for first-stage regressions..44

Table I-7: Robustness Check _OLS...45

Table I-8: Robustness Check _2SLS 46

Table I-9: Robustness Check _Selected Results for first-stage regressions 47

Table II-1: Summary Statistics 89

Table II-2: Baseline Regression Results..90

Table II-3: Selected Results from First-stage Regressions...91

Table II-4: Average of Estimated Coefficients for Monthly Dummy Variables.....................92

Table II-5: Main Regression Results 93

Table II-6: Selected Results from First-stage Regressions with Categorical N.......................94

Table II-7: Average of Estimated Coefficients for Monthly Dummy Variables.....................95

Table II-8: Robustness Check 1: Baseline Regression Results 96

Table II-9: Selected Results from First-stage Regressions...97

Table II-10: Average of Estimated Coefficients for Monthly Dummy Variables...................98

Table II-11: Robustness Check 1: Main Regression Results..99

Table II-12: Selected Results from First-stage Regressions with Categorical N..................100

Table II-13: Average of Estimated Coefficients for Monthly Dummy Variables.................101

Table II-14: Robustness Check 2: Baseline Regression Results 102

Table II-15: Selected Results from First-stage Regressions..103

Table II-16: Average of Estimated Coefficients for Monthly Dummy Variables.................104

Table II-17: Robustness Check 2: Main Regression Results..105

Table II-18: Selected Results from First-stage Regressions with Categorical N..................106

Table II-19: Average of Estimated Coefficients for Monthly Dummy Variables 107
Table III-1: Location of FGD Gypsum Producers and Gypsum Mines 141
Table III-2: Variable Description for EIA and EPA Data 142
Table III-3: Average Annual Summary Statistics 143
Table III-4: Estimates of Plant-level Changes in Natural Log SO_2 Emissions 144
Table III-5: Total Benefits of Reduced SO_2 Emissions Resulting from FGD Gypsum Sales 145
Table III-A1: Tests for Common Trend in SO_2 Emissions 155
Table III-A2: Alternative Specification Tests..156
Table III-A3: First-Stage Estimates of Gypsum Sales Contracts 157

LIST OF FIGURES

Figure III-1: Lime and Limestone Average Price per Metric Ton (1985-2010)....................146
Figure III-2: Gypsum Sales 1985-2010 (Measured in Thousand Metric Tons) 147
Figure III-3: Synthetic Gypsum Market Share 148
Figure III-4: Path Analysis Diagram for Gypsum Sales Contracts and SO_2 Emissions.......149
Figure III-5: Path Analysis Diagram with Unobserved Disposal Regulations......................150

Chapter 1

The Effect of Competition on Price Discrimination in International Flight Market between the U.S. and China

1. Introduction

In this chapter, using a unique self-collected transactional flight tickets dataset from the international flight market between the US and China, I provide new evidence on the relationship between second-degree price discrimination and market competition. Airline pricing is a prominent example of price discrimination. It is a well-known fact that two passengers on the same flight with similar seat positions may have paid very different prices. Economists have long been interested in this phenomenon (e.g. chapter 3 of Tirole 1988; Stole 2007). One of the main issues in the literature is the relationship between the market structure and the ability of the firms to practice price discrimination. The conventional wisdom is that a competitive firm cannot price discriminate because it is a price taker and a firm with market power can price discriminate as long as it can segment the consumers. Hence, as the market becomes more competitive, there should be less price discrimination. However, studies by Katz (1984), Borenstein (1985), Holmes (1989) and Dana (1998) show price discrimination can exist in fairly competitive markets. More recently, Hernandez and Wiggins (2008) group flight tickets into five categories according to ticket qualities, and find that the second-degree price discrimination increases with market competition across low- and high-quality tickets, but decreases with market competition across low- and medium-quality tickets. McAfee, Mialon and Mialon (2006) show the same relationship can be either positive or negative using a third-degree price discrimination model. As theoretical studies do not produce a clear prediction, the relationship between market structure and the degree of price discrimination becomes an empirical question.

The most direct way to examine the empirical relationship between market structure and price discrimination is to collect data on price-quality/quantity schedules and a measure of market competition and estimate the relationship between the variables. Several recent studies in other industries employ this approach. Busse and Rysman (2005) find that in more competitive markets, the relative price of high versus low quality yellow page advertising slots decreases, indicating that more competition leads to less price discrimination. Using the same idea, Clerides and Michis (2006) study the detergent market in six countries and find the relationship is positive in some countries, while negative in others.

Due to data limitations, however, most of the empirical studies in the airline industry use an indirect approach to examine the relationship between price discrimination and market competition. The most popular and widely used dataset in this literature is the DB1A/DB1B ticket and price data from the Department of Transportation's Origin and Destination Survey, which only reports the airfare of the tickets. It does not provide information on ticket restrictions, which can be regarded as measures of the quality of the tickets. As a result, researchers typically examine the relationship between market structure and price dispersion. The rationale is if the main source of price dispersion is caused by price discrimination, then one can infer the relationship between market structure and price discrimination from the relationship between market structure and price dispersion. Applying this approach to a cross sectional data set from DoT's DB1A, Borenstein and Rose (1994) report that as the market becomes more competitive, there is more price dispersion in U.S. airline industry. However, recent panel data studies by Gerardi and Shapiro (2009) and Dai, Liu and Serfes (2010) find the relationship is negative and nonlinear, respectively, using data sets from the same source.

Though these studies are informative, the results from these studies hinge critically on the assumption that one can infer the relationship between market structure and price discrimination from the relationship between market structure and price dispersion. However, price dispersion may come from other sources. For example, Gale and Holmes (1993) and Dana (1999a, 1999b) show that price dispersion can also result from airlines practicing scarcity pricing. The idea is that when demand is uncertain (firms do not know when the peak period is) and holding capacity is costly (as flight service is perishable), airlines may offer more discounted seats during off-peak periods to divert demand away from the peak period. In this case, price dispersion arises because of cost differences and not because of price discrimination. In other words, airlines bear different capacity holding costs for tickets during off-peak versus peak periods, and the price differences between peak and off-peak periods reflect the holding cost differences rather than differences in price discrimination.

Several recent empirical studies also use the direct approach to examine the relationship between market structure and price discrimination in the airline industry. Stavins (2001) collects ticket data for flights within the U.S. from a computer reservation system. The data provides her detailed information on ticket characteristics. She treats tickets with restrictions like advance purchase requirement and Saturday-night stayover as tickets of low quality and examines how the marginal price of these ticket restrictions (price/quality schedule) change as market competition changes. She finds that as the market becomes more competitive, the discount associated with ticket restrictions becomes larger, indicating there is more price discrimination when the market becomes more competitive. Giaume and Guillou (2004) conduct a similar study using data for flights within Europe and reach the same conclusion. Using transacted ticket data from more than 200 routes within the U.S., Hernandez and Wiggins (2008) find mixed results. They report as the market becomes more competitive, the ratio between the price of high-quality tickets and low-quality tickets increases, indicating there is more price discrimination. However, at the same time, they also report the ratio between the price of medium-quality tickets and the price of low-quality tickets becomes lower as there is more market competition.

This paper offers another piece of evidence on the relationship between market competition and price discrimination using the direct approach. Employing a unique dataset on transacted tickets for international flights between the U.S. and China, I examine how the marginal prices of ticket restrictions/qualities change as the market becomes more competitive, as in Stavins (2001), Giaume and Guillou (2004) and Hernandez and Wiggins (2008). My study differs from previous studies in two main aspects. First, previous studies use data from flights within the same continent (the U.S. or Europe), while my data are from international flights between the U.S. and China. The unique aspects of this international market allow me to include some new ticket restrictions/qualities that have previously not been studied and examine how the marginal prices for these restrictions/qualities vary with market competition. Second, previous studies use data from one day (Stavins 2001; Giaume and Guillou 2004) or one quarter (Hernandez and Wiggins 2008) in a mature market, while my data come from several years in a growing market. The panel structure of my data allows me to include route fixed effects in the pricing regression, which control for time invariant unobserved demand and/or supply factors that influence price. These factors have not been controlled in previous studies and doing so makes the pricing regression less likely to suffer from endogeneity bias and allows better identification of the relationship of interest. In addition, during the sample period, the number of daily flights between U.S. and China has almost doubled. This provides additional variation in the market structure variable in the time dimension, which again, allows better identification of the relationship of interest.

Results indicate that the effect of competition on second degree price discrimination is negative. For most of the ticket restrictions/qualities (e.g., Advance, Duration, Weekend, Stops and Direct), marginal prices and hence price discrimination decrease with market competition. The results are consistent with the conventional wisdom and indicate that the firms are less capable in price discriminating when there are more competitors.

The rest of this paper is organized as follows. Section 2 describes the international flight market between U.S. and China. Section 3 introduces the data. Section 4 is devoted to the empirical methodology. Estimation results are discussed in Section 5 and the final Section concludes.

2. The International Flight Market between the United States and the People's Republic of China

The U.S.-China international flight market has expanded significantly during the last three decades due in large part to China's Reform and Opening-up policy of late 1978. The

Civil Aviation Administration of China (CAAC) and Pan American World Airways (Pan Am) began flying between mainland China and the U.S. in the early 1980's,[1] and Northwest Airlines and United Airlines joined the trans-Pacific flight market shortly thereafter.[2] However, since Chinese carriers were not competitive in the international flight market at that time, the Chinese authorities were hesitant to make the market more open to foreign airlines. As a result, the price of a U.S.-China flight ticket was relatively high and demand was low during the 1980's.[3] As time passed, the rapid economic growth of China stimulated huge demand for international flight services between the U.S. and China. Northwest Airlines began service of the first scheduled non-stop flight between the two countries in 1996.[4] More substantial progress of the passenger air service was made in the 1999 Air Services Agreement, which asserted that four airlines from each country were allowed to offer flight services between the U.S. and China. This doubled the weekly number of flights from each country from 27 to 54.[5] Further expansion of the U.S.-China flight market was made in a new agreement in July 2004, which relaxed the restriction that flight services were only allowed between specific cities. Unlimited code-sharing[6] between U.S. and Chinese airline carriers was permitted.[7] Another bilateral agreement reached in July2007, stipulated that by end of 2011, the number of airlines in each country allowed to serve the trans-Pacific market would increase from four to nine and the number of weekly flights would increase

---

[1] Source: http://query.nytimes.com/gst/fullpage.html?res=9903E0DC123BF93AA15752C0A967948260

[2] Source: http://www.everythingpanam.com/.

[3] Source: http://www.voanews.com/tibetan-english/news/a-28-a-2004-07-24-2-1-90264472.html.

[4] Source: Northwest Historical Timeline (1990) from the webpage of Delta Airline.

[5] Source: http://transportation.northwestern.edu/.

[6] Code-sharing refers to an aviation business arrangement where two or more airlines share the same flight. The term "code" generally refers to the two-character IATA airline designator code and flight number. The flight ticket marketing airline may differ from flight operating airline.

[7] Source: http://goliath.ecnext.com/coms2/gi_0199-555172/United-States-China-sign-agreement.html.

from 54 to 249 from each country.[8] Due to the unfavorable market conditions in the international travel industry caused by the U.S. subprime mortgage crisis, the new cap has not been reached as of today. But several new entrants including Delta, US Airways and Hainan Airline (a Chinese airline) started their services in the second half of 2010.

In recent years, there have been numerous studies examining various aspects of the U.S. domestic flight market. Examples include Borenstein (1989), Borenstein and Rose (1994), Dana (1998, 1999a, 1999b) and Gerardi and Shapiro (2009), to name just a few. There have been far less research on international flight markets. Some examples include Brueckner (2001, 2003) and Brueckner and Whalen (2000). Although domestic flight markets and international flight markets are different in some aspects, there are several similarities between the U.S. domestic flight market and international flight markets in terms of marketing flight tickets.

Currently, airlines sell their tickets mainly through two kinds of reservation systems. First, they sell tickets through their own Computer Reservation Systems (CRS's) directly. When a customer calls a ticketing office or visits the website of an airline and purchases a ticket there, the sale is completed using the airline's own CRS directly. The airline does not need to pay any commissions or booking fees to third-party business partners.

Airlines also sell tickets by subscribing to and linking their CRS's to one or several Global Distribution Systems (GDS's) and selling tickets with the help of travel agents. A GDS is essentially a platform that collects ticket information such as fare, schedule and availability of different seats from providers and displays the results globally to travel agents

---

[8] Source: http://www.uschinatrip.com/news.cfm.

who subscribe to it.[9]Examples of GDS's include Sabre in the U. S., Apollo and Amadeus in Europe, Abacus in Japan and TravelSky in China. Airlines subscribe to these GDS's and link their own CRS's to these GDS's to make ticket information available. GDS's are therefore great platforms for airlines to sell their tickets as well as to monitor the pricing of similar tickets by their competitors. Travel agents also subscribe to these GDS's to retrieve ticket information and make reservations for their clients through the systems. The airlines need to pay a subscription fee to the GDS's, and as a result, selling tickets through this channel generates costs to the airlines. In addition, every time a ticket is sold through a GDS, the airline needs to pay a booking fee to the GDS which further increases airline's cost of marketing through GDS's (Orlov 2011).

Although airlines operating in the international passenger air market between the U.S. and China and airlines operating in the domestic U.S. flight market all sell their tickets through CRS's and GDS's, the international U.S.-China flight market is unique in several aspects. One interesting feature is that traditional brick and mortar travel agencies still play a major role in the U.S.-China flight market, while their importance in the U.S. domestic flight market has declined significantly with the arrival of the internet age. Before the arrival of the internet age, selling tickets through GDS's with the help of brick and mortar travel agencies was the dominant channel for airlines to distribute their tickets, as that was the easiest way for the consumers to search for and compare prices from different airlines. Orlov (2011) reports that approximately 80 percent of all airline tickets were sold through brick and mortar travel agencies by the mid-1990's. With the arrival of the internet age, airlines developed

---

[9] GDS's also include information on hotel, car rental and other perishable services.

their own websites and internet-based third-party businesses also emerged (Transportation Group International, 2002).These include online travel agencies (e.g. expedia.com), price search engines (e.g. farechase.com) and bid-based websites (e.g. priceline.com).[10] This made it relatively easy for consumers to search for and compare prices from different airlines by themselves. Also, airlines often offer incentives for consumers to purchase tickets directly from their websites. For example, they offer frequent-flier bonuses when consumers book tickets online (Orlov 2011). As a result, the importance of the brick and mortar travel agencies in the domestic flight market has decreased significantly and nowadays, consumers rarely use brick and mortar travel agencies for the sole purpose of purchasing a domestic flight ticket, though some of them still seek the help from these agencies when they purchase a non-traditional ticket such as an open-jaw ticket, or arrange for a vacation for which a ticket is often bundled with other products like hotel stays and car rentals.

For the international flight market between the U.S. and China, however, brick and mortar travel agencies still play a major role despite the rapid development of internet, due to the following reasons. First, brick and mortar travel agencies help airlines reach potential consumers and serve these consumers better. The majority of the passengers for the international flights between the U.S. and China are Chinese or Chinese Americans who study, work and live in the U.S. and their relatives and visitors. Most of them speak Chinese fluently while some of them are not proficient in speaking English. Therefore, travel agents who speak Chinese can reach and serve these consumers better. Indeed, almost all of the

---

[10] Different internet-based businesses have different sources of ticket information, depending on their business models. Some subscribe to GDS's, some have access to airlines' CRS's through special arrangement and some simply search dozens of websites including the websites of airlines as well as the websites of other online travel agencies.

brick and mortar travel agencies serving this market are run by Chinese or Chinese Americans who know how to speak Chinese and are located in big city Chinatowns.

Second, selling tickets through brick and mortar travel agencies allows airlines to practice price discrimination along a new dimension. Airlines classify tickets into many fare classes and sell tickets in the same class at different prices to practice price discrimination. For example, United-Continental classifies tickets into 21 different fare classes in 7 groups: discounted economy group includes S, T, L, K and G fare classes; economy group includes M, E, U, H, Q, V and W fare classes; discounted full-fare economy group includes fare class B; full-fare group includes fare class Y; discounted business group includes Z and P fare classes; business group includes J, C and D fare classes and first-class group includes F and A fare classes.[11]In the U.S.-China flight market, tickets in certain fare classes, usually those tickets with heavy discounts, are only available through GDS's during a specific time period (usually two or three months before the departure date during peak time and one month before the departure date during off-peak time). Each airline usually grants access of such tickets to just a few consolidators (a special type of travel agents who specialize in wholesaling tickets to other travel agents) using GDS's.[12] The consolidators then re-sell the tickets to brick and mortar travel agencies in Chinatowns (by passing the access rights to the heavily discounted tickets in GDS's), who then re-sell the tickets to consumers.[13] The heavily discounted tickets are not available from the websites of the airlines or through any other

---

[11] Source: http://www.united.com/CMS/en-US/Marketing/CustComm/Promotions/Pages/united.aspx.

[12] Consolidators play an important role in international flight markets, but only a minor role in the U.S. domestic flight market. To learn more about consolidators, see the article at http://tourism.about.com/od/travelagentsandagencies/a/Travel-Agents-Utilizing-Airfare-Consolidators.htm.

[13] For more information about how this process works, see the article at http://www.smartertravel.com/travel-advice/the-truth-about-consolidator-fares.html?id=9581452.

agencies (either online or brick and mortar) that do not work with the consolidators. In this way, the airlines essentially segment the consumers into two groups. One group of consumers (mainly Chinese or Chinese Americans) knows cheaper tickets are available through brick and mortar travel agencies in Chinatowns, while the other group (mainly people who are not Chinese or Chinese Americans) does not know about this supply source, although the agents do not intentionally block the non-Chinese consumers out.[14] The second group of consumers therefore purchases more expensive tickets from other sources.

3. Data

3.1 Ticket Data

The data used in my analysis come from several sources. The main components of the data, the ticket and price data, were collected from the website http://www.travelsuperlink.com/. This website started in 2005 as a simple platform that allows air travelers between the U.S. and China to share with one another their recent purchasing experiences, with the goal to help travelers find good deals. To contribute to this effort, a traveler can upload the information of the ticket that was purchased through a ready-to-fill form on the website. The information collected includes the purchase date, the departure date, the return date (if it is a round trip), the trip type (round/single/open/open-jaw),[15] the departure city, the arrival city, the connecting cities if there are any, the airfare

---

[14] The agents speak English in general, and they do serve every client who can access to them, although the consumers with Chinese background are normally expected by them.

[15] Both open and open-jaw tickets are for round trips. An open ticket is a ticket whose return date is flexible. A ticket is open-jaw if the traveler's final destination of the round trip is different from the departure city. For example, the round trip ORD-PEK, PEK-JFK is an open-jaw ticket. It can also be the case that the traveler returns from one city other than the one he/she arrived. For example, the round trip ORD-PEK, PVG-ORD is an open-jaw ticket, too.

amount, the airline carrier, the name of the travel agency through which the ticket is booked, and any comments regarding the quality of the service provided by the agent. Nowadays, it is not only a place where travelers share their information, but also a place where ticket agents and other businesses in the travel industry advertise their services and special deals. The information a user can get from the website is no longer confined to the U.S.-China air ticket deals, but also includes travel information with destinations to almost all parts of the world and through other transportation modes like cruises and car rentals.

My analysis is based on the reported transactions on this website for the period from1/1/2005 to 12/31/2011. A ticket was included in the dataset if both its departure date and return date are between these two dates, it is a round-trip economy class ticket and the destination/origin city in China is either Beijing or Shanghai.[16] In total, there were 16,397 such tickets from 515 routes reported on the website during the sample period. After deleting observations with missing or outlier values, 15,232 tickets from 468 routes were left.[17] A route is defined as a triplet of origin-destination-origin airports. Therefore, PEK-ORD-PEK and ORD-PEK-ORD were treated as two different routes. Using the reported information I created the following variables for each ticket: the airfare amount (Price), whether the trip originates from the U.S. (USorigin), whether the trip is direct (non-stop) or not (Direct), whether the origin or the destination city is one of the hubs of the air carrier (Hub), dummies

---

[16] Including tickets to or from other cities in China in the analysis requires data on flight schedules in China, to which I do not have access.

[17] More specifically, 740 were deleted because the carrier information is missing; 250 were deleted because the reported purchase date is after the departure or return dates or the reported purchase date is 11 months before the departure or return dates (flight tickets are usually available for purchase up to 11 months in advance); 9 were deleted because of duplicate entries for the same ticket; 83 were deleted because the reported return date is earlier than the departure date; 55 were deleted because the origin and destination airports are the same and 28 were deleted because the ticket price was below $400 or above $3,000.

for air carriers, number of days between the purchase date and the departure date (Advance), number of days between the departure date and the return date (Duration), whether the origin or the destination city in China is Beijing rather than Shanghai (PEK), monthly dummies for the US to China leg of the trip, monthly dummies for the China to US leg of the trip, route dummies and the number of flights dates that are during peak period (Peak) and weekend (Weekend). In this market, weekend is defined to be Friday, Saturday, Sunday and Monday, not just Saturday and Sunday. This is because connecting flights (flights within the U.S.) on these four days have higher load factor. Connecting flights on Mondays and Fridays have higher load factor because more domestic travelers travel on these two days. Connecting flights on Saturdays and Sundays also have higher load factor because fewer connecting flights are scheduled on Saturdays and Sundays. As data indicated, peak travel period is defined as travel in January, May, June, July, August and December, which correspond to winter and summer vacation time.

3.2 Official Airline Guide (OAG) Data

The second data source for my empirical analysis is the worldwide issues of the Official Airline Guide (OAG), which report schedule information for almost all commercial flights in the entire world. For each ticket, using the reported information on departure date, return date, origin airport and destination airport, I first searched the OAG databases to identify all the feasible itineraries by all airlines that correspond to the ticket. A feasible itinerary is defined as an itinerary that has the same departure and return dates, the same origin and destination airports, a total travel time less than 30 hours, up to 2 stops within U.S., up to 1 stop in China and up to 1 stop in a third country like Canada, Japan and Korea, a layover time more than 60 minutes in the connecting airport where passengers need to go through border control and a layover time more than 40 minutes in the connecting airport(s) where passengers do not need to go through border control. Note that for the same ticket, a feasible itinerary can differ from the reported itinerary in terms of the carrier and/or the connecting cities. There are 14,320 out of 15,232 observations that can be matched with feasible itineraries.

I then compared the feasible itineraries with the reported itinerary for each ticket to verify whether the itinerary information reported by the consumers is accurate or not. Based upon the matching degree between the reporting itinerary from price data and the corresponding OAG itinerary, the 13,641 observations can be further classified into several groups[18]. There are 10,344 observations that are perfectly or partially but one-to-one[19] matched with OAG schedule data. Another 3,297 observations can be partially, but one-to-many matched with schedule data. Robustness check is conducted by including those observations for regression.

Then, using the confirmed itinerary information, I created the following variables for each ticket: total distance travelled (Distance), total layover time (Wait), number of extra

---

[18] Another 679 observations of the 14,320 ticket sample cannot be matched with the OAG schedule data at all, based upon the connection(s) information. Those observations are not included in any analysis due to the fact that the corresponding variables, such as, Wait and Stops cannot be created as describe below.

[19] One-to-one perfect match refers to the observations that one price observation can be perfectly matched and only matched with one schedule observation generated from OAG data in regards to itinerary; one-to-many partial match refers to the price observations that one price observation can be match with more than one schedule observations, and the connection information reported by price observations can be only partially matched with the schedule observations. The schedule among the matched itineraries with the shortest distance is kept for further variable creations.

stops[20] (Stops) and whether any of the origin, destination and connecting airports have slot controls (Slot).[21] More importantly, using the feasible schedules data generated from OAG, I created the variable measuring market competition. The market competition variable is the key explanatory variables in my regression analysis. Previous studies used market shares on a route, either in terms of the number of flights operated or in terms of the number of passengers transported, as measures for market competition. For example, Stavins (2001) used the Herfindahl-Hirschman index (HHI) in terms of the number of flights. Lacking data on the actual number of passengers transported by each airline for each route,[22] I use the number of other airlines (Comp) that can serve the same route on the same departure and return dates as the measure for market competition.[23] This variable should be highly correlated with a market competition variable based on the number of flights operated. This is because for most of the routes, each airline only has at most one daily flight for the international segment available to serve the consumers.[24]Some international carriers did not fly every single day of a week, and some of the code share contracts among carriers changed over time. I took these into account when creating this variable.

---

[20] The number of extra stops is the difference between the number of stops for the ticket's itinerary and that of the feasible itinerary with the smallest number of stops.

[21] According to Czerny (2008), slots at DCA and LGA were regulated throughout my sample period, while JFK was regulated until January of 2008.

[22] For example, Continental serves the route RDU-PEK-RDU with a connection at EWR. Though data on the number of passengers transported by Continental between RDU and EWR is available from BTS, I do not know how many of them are consumers heading to PEK.

[23] Some of the previous literature (Stavins 2001, Hernandez and Wiggins 2008) used Herfindahl–Hirschman Index, or HHI to measure market competition. HHI is constructed out of market share. This analysis is not able to create market share variable due to the fact that 77% of the observations have connecting flights. For example, RDU-ORD-PEK, the market share of passengers for the segment of RDU-ORD can be created, but it is not possible to check the percentage of passengers fly from RDU to ORD and continually fly towards PEK.

[24] This is different from the case of trans-Atlantic market. For example, both American Airlines and United Airlines offer several daily flights between New York and London.

3.3 Bureau of Transportation Statistics (BTS) Data

My third data source is several datasets from the Bureau of Transportation Statistics (BTS). First, the Data Bank 28DS and 28IS datasets provide information on the number of passengers transported and the number seats available on non-stop flights between a pair of airports by airline and month. Data Bank 28DS covers airport pairs where both airports are located in the U.S., while Data Bank 28IS covers airport pairs where one airport is located in the U.S. and the other airport is located outside of the U.S.. For each observation in my ticket data, I first used the BTS data to compute the load factor for each segment of the itinerary. The load factor was defined as the ratio between the total number of passengers transported and the total number of seats available from all flights by the corresponding airline in the corresponding month. The flights included both the flights operated by the airline itself and flights operated by its code share partners. I then computed the weighted average load factor variable (Load) for the ticket by averaging across all the load factors for its segments, using the segment distance as the weight. I was able to create the Load variable for 14,004 out of the 14,320 tickets that survived the data cleaning procedures described above. For 260 tickets, the Load variable could not be computed because the itinerary involves a stop in a third country and as a result, the load factor for the segment between China and the third country could not be computed because the BTS datasets do not cover such flights. For the other 56 tickets, the Load variable was not created simply because no corresponding statistics could be found in the BTS data for some or all of the segments in the itinerary.

Then, I used the Origin and Destination Survey-DB1B dataset to create the airport dominance (Domo) variable. The DB1B dataset provides quarterly information on the number of passenger originations by each domestic airline at each airport in the U.S..[25] For each observation in my ticket data, I first used the information from the DB1B to compute the airline's shares of passenger originations for all the domestic airports on the itinerary during the corresponding quarter. Then, I took the average of shares across all the airports as the Domo variable. For 1,075 tickets out of the 14,004 tickets survived from the data cleaning procedure up to this point, the airline is a foreign airline and does not report statistics to BTS. I set the Domo variable for these tickets to be zero as foreign airlines (especially those airlines which serve the US-China market) usually account for a small share of passenger originations at U.S. airports.

3.4 Census Data

Finally, I obtained the population and per capita income data from the Census. For each ticket, I got the number of total population (Population), the number of Chinese-ethnicity (except Taiwan origin) population (Chinese) and per capita income (Income) from the corresponding metropolitan and micropolitan statistical areas the U.S. origin/ending airport on the itinerary serves.[26,27]For each observation, all these data are from the year when the ticket was purchased.

---

[25] An origination is the beginning of a directional trip.

[26] Some airports serve multiple MSAs. For example, airport MBS, serves Midland, MI, Micro Area, Bay City, MI Metro Area and Saginaw-Saginaw Township North, MI Metro Area.

[27] Though each ticket has two ending cities, one in China and one in U.S., I did not average the population and income variables across the two ending cities. This is because Beijing and Shanghai have similar population and per capita income. Therefore, taking averages across the two ending cities would yield variables that have similar cross-observation variations as these variables for the U.S. city alone.

3.5 Summary Statistics

Table 1 provides the summary statistics for the variables that used in the estimation. Several features of the data are worth mentioning. First, there is significant variation in the observed price. The prices for the observed tickets range from $442 to $2,966, with an average of $1,109. Hence, the range is about 228% of the mean. The large variation can be explained by both demand and supply side effects. The data include both tickets for flights from or to an airport on the west coast and tickets for flights from or to an airport on the east coast. If it were the demand side effects, it means that fewer passengers travel between the west coast and China, which is contrary to the fact that more people with Chinese background live at the west coast of America. Then part of the variation comes from the supply side: the west coast is closer to China and hence the price is lower. The remaining variation comes from other sources, including the price discrimination practice by the firms. The range of the price variable also indicates that all of the observed tickets are of economy class.[28] Second, since a large variety of routes are involved for the transacted tickets and the data spans for six years, there are significant variations in the market competition variable, which has an average of 6 and a range of 12. Third, there are significant variations in the ticket quality/restriction variables (Advance, Duration, Peak, Weekend, Wait, Stops, and Direct), of which the marginal prices are taken as proxies for price discrimination measures. The large variations in these variables help identify the impact of market competition on price discrimination.

---

[28] Business class tickets cost at least $3,000 and first class tickets cost at least $10,000 during the sample period.

4. Empirical Models

To examine the relationship between price discrimination and market competition, I regress the transacted ticket price on ticket restrictions/qualities, the market competition measure, their interaction terms, as well as other route-, airline- and ticket-specific factors. The empirical model is specified as follows:

(1). $$Price_{ijkt} = \beta_0 + \beta_1 R_i + \beta_2 Comp_{ikt} + \beta_3 R_i * Comp_{ikt} + \beta_4 X_{ijkt} + \varepsilon_{ijkt},$$

where $Price_{ijkt}$ is the price of ticket i by airline j on route k and dates t.[29] R_i denotes the ticket restriction/quality variables, which include $Advance_i$, $Duration_i$, $Weekend_i$, $Peak_i$, $Wait_i$, $Stops_i$, and $Direct_i$. In my data, I do not observe whether the ticket comes with an advance purchase requirement as in Stavins (2001) or when the cutoff date for the advance purchase requirement is. However, I observe the $Advance$ variable, the number of days in advance the ticket was purchased, which should be able to capture the effect of the advance purchase restriction on ticket price. $Duration$ captures the stay restriction associated with the ticket. In the U.S. domestic market, airlines offer discounts to tickets with a Saturday-night stayover requirement, as studied by both Stavins (2001) and Hernandez and Wiggins (2008). In the international market between the U.S. and China, stay restriction is also an effective tool for the airlines to segment consumers as some consumers are willing and able to stay longer than others. $Peak$ and $Weekend$ indicate travel restrictions associated with the ticket. Airlines often offer discounted tickets with the restriction that the tickets cannot be used for travel on certain days of a week or during certain time period. The travel

---

[29] Each t corresponds to two dates, one for the US-China leg and the other for the China-US leg.

restriction studied by Hernandez and Wiggins (2008) is similar to my $Weekend$ variable here. $Wait$, $Stops$ and $Direct$ are ticket quality variables. A trip with more stops and longer layover time is less convenient to the consumer and the consumer is compensated with a lower price. To the best of my knowledge, no study has ever studied the price discrimination with respect to $Peak$, $Wait$, $Stops$ and $Direct$.

$Comp_{ikt}$ is the measure for market competition discussed in section 3.2 above. X_{ijkt} collects other control variables that are likely to influence the ticket price. It includes Hub_{jkt}, $Distance_{ik}$, $Distance_{ik}^2$, $Slot_{ikt}$, $Load_{ijkt}$, $Domo_{ijkt}$, $Income_{kt}$, $Chinese_{kt}$, airline and route dummies.[30] In addition, X_{ijkt} includes two sets of monthly dummies (one for each of the 72 months of the sample period) for flight dates. Since all the tickets in my analysis are of round-trip, each ticket has both a US-China leg and a China-US leg. As seasonality and time trend are important determinants of ticket price, I use two sets of monthly dummies to control for them, one set for when the US-China leg took place and the other set for when the China-US leg took place. For both sets, the dummy variable for the first month of year 2005 is the omitted month. In addition, because of the inclusion of monthly dummies, the coefficient for the $Peak$ variable in β_1 is not separately identified due to perfect multicollinearity and is therefore omitted.

---

[30] Including both the distance and route dummies in the regression does not cause multicollinearity problem because two tickets on the same route can still have different travel distances due to different connecting airports.

4.1 Identification

There are two identification issues in estimating and interpreting the coefficients in (1): one is conceptual and the other is econometric. To illustrate the first issue, I use the ticket restriction variable $Advance$ as an example.[31] Suppose there are two tickets that are identical in all aspects (in (1), this is achieved by including other variables in the regression) except that the first ticket is purchased d_1 days in advance and the second ticket is purchased d_2 days in advance with $d_1 > d_2$. Let p_1 and p_2 denote the prices for the two tickets and c_1 and c_2 the marginal costs of the two tickets. The price difference of the two tickets can be decomposed into two parts, as follows:

(2). $$p_1 - p_2 = (p_1 - c_1) - (p_2 - c_2) + (c_1 - c_2).$$

In equation (2), the first two terms in parenthesis measure price discrimination, which is defined as the difference in price-cost markups, and the last term measures the cost difference. The estimated coefficient on $Advance$ (β_1) in equation (1) provides an estimate for $p_1 - p_2$. Therefore, β_1 can only be interpreted as a measure for price discrimination under the assumption that $c_1 = c_2$. As Dana (1998) points out, the marginal cost for a ticket in the airline industry is better described as the sum of the marginal cost of production and a shadow cost of capacity. The marginal cost of production is incurred only when the ticket is sold (e.g., the cost of serving meals and drinks), but the shadow cost of capacity is incurred irrespective of whether the ticket is actually sold. In the above example, for the two tickets that differ only in purchasing date, the marginal production costs are likely to be the same.

---

[31] Though not discussed at length as I do here, both Stavins (2001) and Hernandez and Wiggins (2008) employ this assumption, either implicitly or explicitly.

However, their shadow costs of capacity can be quite different. At the later purchase date, fewer seats are left on the flight and the opportunity cost of each additional seat is higher. Therefore, the shadow cost of capacity for the second ticket is likely to be higher and hence c_2 is likely to be higher than c_1.

Although β_1 might be a biased estimate of price discrimination if $c_1 \neq c_2$, for my purpose, it is the sign of β_1 that matters. In this study, the focus is on the relationship between market competition and price discrimination. If β_1 and β_3 in equation (1) have the same signs, then more competition leads to more price discrimination. On the other hand, if β_1 and β_3 have opposite signs, then more competition leads to less price discrimination. Therefore, as long as $(p_1 - c_1) - (p_2 - c_2)$ in equation (2) has the same sign as $c_1 - c_2$ or has the opposite sign but dominates $c_1 - c_2$ in terms of absolute values, $p_1 - p_2$ in equation (2) has the same sign as $(p_1 - c_1) - (p_2 - c_2)$ and the sign of the estimated β_1 has the same sign as that of price discrimination. So the identification assumption here is that the difference in mark-ups has the same sign as the cost difference or has the opposite sign but dominate the cost difference in terms of absolute values. Stated alternatively, it means that the relatively costly ticket also has a higher mark-up. This assumption is weaker than the assumption $c_1 = c_2$.

The identification assumption for β_3 is similar. If when the market becomes more competitive, only the difference in mark-ups changes then β_3 captures how price discrimination changes as market condition changes. Again, the difference in the marginal production costs of the two tickets with different purchasing dates is not likely to change when market competition condition changes as the difference is close to 0 in all markets. But the difference in the shadow costs of capacity for the two tickets might change when the market becomes more competitive. In a more competitive market, there are more flights available and hence the shadow costs of both tickets are low. In a less competitive market, there are fewer flights available and hence the shadow costs of both tickets are high. If the ratios between the two shadow costs are the same across the two markets, then the difference in shadow costs in the less competitive market is higher than that of the more competitive market. In this case, the shadow cost difference and hence $c_1 - c_2$ increases when the market becomes more competitive. As a result, β_3 from equation (1) is a biased estimate of the relationship between market competition and price discrimination. But again, when the market becomes more competitive, as long as the difference in mark-ups changes in the same direction as the difference in marginal cost, or changes in the opposite direction but dominates the change in marginal cost in absolute terms, then the sign of β_3 estimated from equation (1) is still correct.

The second identification issue is purely an econometric one. In estimating a pricing function like the one specified in equation (1), it is well known that the market competition variable, $Comp_{ijkt}$, might be endogenous, that is, $Comp_{ijkt}$ might be correlated with the error term ε_{ijkt}. Any uncontrolled time varying[32]route or airline specific demand and/or supply

---

[32] Time-invariant unobserved route specific demand and supply factors have been controlled by the route dummies.

factors that influence the ticket price and the number of airlines serving the route at the same time can cause such an endogeneity problem.

In the particular context of this study, however, the $Comp_{ijkt}$ variable is likely to be exogenous after controlling for many ticket price determinants including the full set of route dummies. Suppose route k is a route between a non-gateway airport in the US and PEK in China.[33] Whether airline j is able to serve this route depends on whether the airline offers flights between the non-gateway airport and one of its gateway airports, which are usually hubs of the airline, and whether the airline offers flights to China, which, by definition, depart from one of its gateway airports. Both of these decisions are not likely to be determined by time varying route k specific demand and supply conditions. First, whether the airline offers flights between the non-gateway airport and one of its gateway airports or hubs (and hence includes the non-gateway airport as part of its network) depends on the demand for flight services by passengers originating from (going to) this airport to (from) all airports on the airline's network, not just by passengers for route k. Similarly, whether the airline offers flights from gateway airports to China depends on the demand for flight services by passengers to (from) China from (to) all its airports in the US, not just passengers for route k. Therefore, the number of airlines that are able to serve route k is not likely to be correlated with time varying route k specific demand and/or supply conditions.

In order to address any potential endogeneity issue, an instrumental variable (IV) approach is used. In the context of this study, a valid IV needs to satisfy two conditions.

---

[33] Gateway airports are airports with international flights.

First, it needs to be correlated with the potentially endogenous variable, that is, $Comp_{ijkt}$. Second, it needs to be uncorrelated with the uncontrolled time varying route k specific demand and supply conditions. I use the population of the US end MSA of route k in the year when the ticket was purchased (Pop_{kt}) as the IV variable for $Comp_{ijkt}$. Borenstein and Rose (1994), Hernandez and Wiggins (2008), Gerardi and Shapiro (2009) also use popultation as an IV for market competition variables in their studies. As discussed above, whether an airline is able to serve route k depends on whether the airline offers flights between the local airport and one of its gateway (hub) airports. Airlines are more likely to offer flights from their hubs to airports with more population as more people demand more flight services. Therefore, Pop_{kt} is likely to be correlated with $Comp_{ijkt}$. On the other hand, population is not likely to be correlated with the uncontrolled time varying demand and/or supply factors for flight services to China, especially after I have already controlled for the number of Chinese-ethnic population in (1).

5. Results

Based upon the matching degree between the reporting itinerary from price data and the corresponding OAG itinerary, the 13,222 out of the 14,004 observations can be further classified into several groups[34]. There are 10,038 observations that are perfectly or partially but one-to-one[35] matched with OAG schedule data. Those observations are used as the main

---

[34] The rest of the 14,004 ticket sample (782) cannot be matched with the OAG schedule data at all, based upon the connection(s) information. Those observations are not included in any analysis due to the fact that the corresponding variables, such as, Wait and Stops cannot be created.

[35] One-to-one perfect match refers to the observations that one price observation can be perfectly matched and only matched with one schedule observation generated from OAG data in regards to itinerary; one-to-many partial match refers to the price observations that one price observation can be match with more than one

analysis. Another 3,184 observations can be partially, but one-to-many matched with schedule data. A robustness check is conducted by including those observations for regression.

5.1 Baseline Results

I first estimate a restricted version of (1) with the coefficients for the ticket restriction/quality by market competition interaction variables, β_3, set equal to zero. Also, since the relationship between ticket price, $Advance$, and $Duration$ are likely to be nonlinear and complex, I have used a set of categorical dummy variables which can be regarded as a non-parametric specification, rather than a parametric specification, for these two variables. The following categories are used for the $Advance$ variable: 0-7 days, 8-14 days, 15-21 days, 22-28 days, 29-35 days, 36-45 days, 46-60 days, 61-90 days, 91-120 days, 121-150 days, 151_180 days and over 181 days. For the $Duration$ variable, the following categories are used: 0-5 days, 6-13 days, 14-21 days, 22-30 days, 31-45 days, 46-60 days, 61-92 days, 93-120 days, 121-183 days and over 184 days.

Results are collected in Table 2. The left panel reports the OLS regression results. The estimated coefficients for most of the variables have the expected signs. Here I only discuss those estimates that are statistically significant. First, regarding the $Advance$ variable, in general the ticket price is lower if the ticket is purchased more than 7 days in advance compared with less than or equal to 7 days in advance (the omitted $Advance$ category in the regression). The only exception is if the ticket is purchased more than 180

---

schedule observations, and the connection information reported by price observations can be only partially matched with the schedule observations. The schedule among the matched itineraries with the shortest distance is kept for further variable creations.

days in advance. In that case, the ticket price is actually higher. These results are consistent with my prior expectations. People who are able to purchase the tickets early usually have a better idea of when they can travel between the U.S. and China. These people are usually students and retired people. They have lower income and hence a more elastic demand. People who cannot book tickets early are usually people who do not have a good idea on when they can travel. These people include some fully employed workers with a tight work schedule or people who travel for emergencies. They usually have higher income and/or a less elastic demand. As a result, airlines offer people who buy early a lower price compared with people who buy less than or equal to 7 days in advance. Also, the results indicate that the relationship between ticket price and *Advance* is quite complex and nonlinear. Consumers who buy 90 to 120 days in advance pay the lowest price, while consumers who purchase more than 180 days pay the highest price. The latter result is probably due to the fact that airlines usually start offering discounts when it is within 5 or 6 months of the departure date. As a result, consumers who purchase more than 180 days in advance are more likely to pay the full ticket price.

The *Duration* variable tells a similar story. Again, the relationship between ticket price and *Duration* is nonlinear. In general, the ticket price is lower if the consumer stays longer than 5 days, compared with a stay less than or equal to 5 days (the omitted *Duration* category in the regression). But only people who stay between 22 days and 60 days pay a significantly lower price than people who stay less than or equal to 5 days. People who stay more than 5 days and less than 22 days and people who stay more than 60 days do not pay a statistically different price than people who stay less than or equal to 5 days. In this market, different kinds of consumers usually stay for different days. People who can only stay for no more than 5 days are usually fully employed people with a tight work schedule or people who are making emergency travels. These people tend to be high income people and/or demand inelastic. Tourists usually stay for 1 week or 2 weeks. Students usually stay for one month or two months, depending on whether it is winter vacation or summer vacation and retired people usually stay for more than 3 months. Students are the group with the lowest income and hence have the most elastic demand, which explains why airlines offer a significantly lower price for people who stay between 22 and 60 days.

Regarding other ticket restriction/quality variables, the ticket price is higher if the flight is fulfilled during weekends or peak period.[36] Table 3 reports the averages of the coefficients for the monthly dummies by year and by whether it is during the peak travel period or not. It is clear that tickets for travel during the peak travel period command a higher price. The wait time has a surprisingly positive effect to airfare. It may be inconvenient or uncomfortable for some consumers to wait at the airport for the next flight, however, for most of the international travelers, abundant transferring time for passing through customs and transferring to the next flight gate is preferred. The number of stops represents the inconvenience level of the flight, and it has a negative impact on the ticket price as expected.

Second, competition has a negative effect on price, as expected. One more competitor on the same route decreases the price for about $3.6. Third, also as expected, the ticket price is higher if the airline plays a dominant role at the airport.

---

[36] The estimated coefficients (not reported but available upon request) for year-month dummies indicate that ticket prices are significantly higher during summer and winter, the peak travel period.

Fourth, compared with United Airlines, whose dummy variable is omitted in the regression, most of the other airlines charge lower prices. This is not surprising as UA has been the dominant player in this market and has the largest number of daily direct flights between the U.S. and China. Table 1 shows 32% of the tickets in my dataset are tickets sold by UA. As a dominant player, UA has more loyal customers it captures through its frequent flyer program over the years and other airlines have to deeply discount to attract customers from UA. The cost of switching to another airline (even just temporarily) is higher in the international flight market than in the domestic one. A U.S.-China roundtrip usually results in an increase of about 14,000 miles in one's frequent flyer's account. Therefore, buying a ticket from another airline usually means giving up the opportunity to boost one's frequent flyer's account and significantly slowing the progress to win a reward ticket. Fifth, the average personal income increase by $10,000, the ticket price decreases by $11, and the result is statistically significant. Lower price indicates a higher supply or a lower demand, in the context of this market, the explanation should come from the supply side. It means that wealthier metro/micropolitan areas enjoy abundant of flight services, and therefore also lower ticket price.

Next, I run the 2SLS regression, using population as the instrument for the total number of competitors variable. Results are collected in the right panel of Table 2. The bottom of Table 2 shows that the Angrist-Pischke (2009) χ^2 test rejects the hypothesis that the IV regression is under identified and the Angrist-Pischke (2009) F test rejects the hypothesis that the IV regression is weakly identified. Regarding coefficient estimates, the IV regression yields very similar results to those of OLS, both in terms of parameter values and statistical significances. One of the exceptions is the coefficient for the market competition variable. The IV regression yields a much larger estimated impact of competition on ticket price. Each additional competitor leads to a reduction of $70 in ticket price, rather than just $3.6, as from the OLS estimation, although the result is not statistically significant. If the error term in (1) captures some demand and/or supply factors that cause ticket price and the number of competitors in a market to move in the same direction, then the OLS estimate is likely to be upward biased. The IV regression corrects this bias. The other exception is that, compared with United Airlines, most of the rest of airlines still charge a lower price, but the results are not statistically significant anymore. In contrast, Chinese airlines charge a higher price than UA does and the result is significant. Although the Chinese airlines cannot compete with the United Airlines in regards to the flight frequency or the market shares, those airlines are relatively more popular among the Chinese consumers, and therefore, the loyal consumers pay relatively higher price.

5.2 Main Results

I then estimate the unrestricted specification of (1), or the specification including the ticket restriction/quality-market competition interaction terms. Results from the OLS estimation are collected in Table 4, the results from the IV regression are collected in Table 5 and some of the selected results for first-stage regressions are collected in Table 6. As Table 6 shows, the Angrist-Pischke χ^2 test rejects the hypothesis that the IV regression is under identified and the Angrist-Pischke F test rejects the hypothesis that the IV regression is weakly identified. In the IV regression, the ticket restriction/quality-population interaction terms are used as IVs for the potentially endogenous ticket restriction/quality-market competition interaction terms. Again, the estimation results from the OLS and the IV regressions are fairly similar, both in terms of parameter values and statistical significances. Below, I focus my discussion on the OLS results.

Comparing with the OLS results in Table 2 that excludes the interaction terms, more coefficients for the *Duration* categorical dummies are significant and the estimated parameter values for these coefficients are larger in terms of absolute values. Estimates for other coefficients are similar and the adjusted R^2 only improves slightly.

The parameters of interest here are β_1 and β_3 in (1). As discussed in section 4.1 above, if the coefficient for the ticket restriction/quality variable and the coefficient for the corresponding ticket restriction/quality-market competition interaction term have the same signs, then this indicates price discrimination increases as the market becomes more competitive. On the other hand, if the two coefficients have the opposite signs, then this indicates price discrimination decreases with market competition.

For the *Advance* categorical dummies variables, the interaction terms are not statistically significant other than the category with over-181 days purchase in advance. The coefficient for the Adv181_over dummy variable and the coefficient for the corresponding interaction term have the opposite signs, indicating price discrimination with respect to the number of purchasing days in advance decreases with market competition.

Results from the *Duration* categorical dummy variables tell a similar story. All of the coefficients for the *Duration*-market competition interaction terms are statistically significant. And for each category, the coefficient for the dummy variable and the coefficient for the corresponding interaction term have the opposite signs, indicating price discrimination with respect to the number of days in stay also decreases with market competition. The magnitude of the effect of market competition on price discrimination is non-trivial as well. For example, compared to a consumer who only stays for less than 6 days, a consumer who stays between 6 and 13 days pays $225 less for the otherwise same ticket in a market where there is no competitor to the airline from which the consumer is purchasing the ticket. For each additional competitor in the market, the expected discount is $32 less. The result for *Weekend* variable is significant, and consistent with the variables discussed above. The coefficient for the *Weekend* variable and the coefficient for the corresponding interaction term have the opposite signs. It means that price discrimination with respect to the *Weekend* decreases as the market competition increases.

Results from the *Peak* variable, however, show a different story. The coefficients for the monthly dummies show that tickets for travel during the peak period are significantly higher.[37] The coefficient for the peak-market competition variable is also positive and statistically significant. This indicates that the price discrimination with respect to travel during the peak period increases with market competition, with one more competitor leads to an increase of about $2.4 in the price premium paid for traveling during the peak period. The result is not economically significant though.

5.3 Robustness Check

As mentioned above, there are 3,184 price observations can be partially and one-to-many matched with schedule observations in the sample. In order to check whether the results are consistent, we repeat all of the regressions with a bigger sample where the 3,184

---

[37] Estimates not reported in Tables 4 and 5 for brevity, but are available upon request.

price observations are included. The 3,184 price observations can be partially matched with the corresponding schedule observations, which means that the information reported by the consumers are mainly trustworthy, and the connection information cannot be perfectly matched with schedule information out of the possibility that the consumers skipped part of the connection information report for simplicity. Since the matching schedule is not unique, I picked the ones that match the itinerary the most and cover the shortest distance. First of all, the schedule information that matches the reported itinerary the most have the least number of connections, and therefore the distance will be the shortest among all of the matching observations[38]. Second of all, even if there are multiple matching itineraries that have the same number of stop(s) but different connection airport(s), the difference in distances across the itineraries is trivial[39]. Results are reported in Tables 7-9, and they remain qualitatively the same and quantitatively similar.

In summary, my results show that there is a negative relationship between second degree price discrimination and market competition. This means that the ability for airlines to practice price discrimination in regards to ticket quality/ticket restriction is deteriorated as there are more competitors, and therefore, the price dispersion across the tickets within a consumer group declines when consumers have more options to choose from in regards to service suppliers. This is consistent with the conventional wisdom. Among the ticket

---

[38] For example, if a reporting ticket shows the itinerary as RDU-PEK, without connection information, the possible matching schedules, however, all have connection information, for example, RDU-DTW-PEK and RDU-DTW-NRT-PEK; the RDU-DTW-PEK is picked over the RDU-DTW-NRT-PEK.

[39] For example, if the reporting ticket shows the itinerary as PGV-CLT-PEK, but the possible matching schedules all have another connection in the US, for example: PGV-CLT-ORD-PEK (about 7,380 miles one way) and PGV-IAD-ORD-PEK (about 7,443 miles one way), the one with least distance (PGV-CLT-ORD-PEK) will be picked. The difference between the two itineraries is less than 1% of the average.

restriction variables, Advance, Duration, Weekend tell the same story significantly, the results from the other variables, Wait, Stops are consistent, though insignificant. There is only one “outlier”, Peak, tells the opposite situation. But the result there is not economically significant. Furthermore, in peak load pricing, higher price during peak time is used as the method to divert the demand instead a price discrimination tool. The result from Peak variable probably is not supposed to be included in the indication between price discrimination and market competition.

6. Conclusions

In this chapter, using a unique dataset from the US-China international flight market, I offer new evidence on the relationship between second-degree price discrimination and market competition in the airline industry. Results show that as the market becomes more competitive, airlines price discriminate less with respect to the ticket restrictions/qualities. The IV regressions and OLS regressions yield similar results, in terms of both economical and statistical significances.

As discussed in section 4.1 above, my results depend on the identification assumption regarding the relationship between the difference in mark-ups and the difference in marginal costs for tickets with different restrictions/qualities. Although the identification assumption cannot be tested, it is a weak assumption. When airlines are not able to segment consumers explicitly, they offer a menu of price-ticket quality/restriction and let consumers self-select, and practice second-degree price discrimination. It will be interesting to further examine the relationship between the price dispersion across consumer groups and market competition under the scenario that airlines are able the segment consumers into different groups. This will be explored in Chapter 2 of this dissertation.

REFERENCES

Angrist, J. D. and J.-S. Pischke (2009): Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

Borenstein, S. (1989): "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry,"*RAND Journal of Economics*, 20, 3, 344-365.

Borenstein, S. (1985): "Price Discrimination in Free-Entry Markets,"*RAND Journal of Economics*, 16, 3, 380-397.

Borenstein, S. andN. Rose (1994): "Competition and Price Dispersion in the U.S. Airline Industry,"*Journal of Political Economy,* 102, 4, 653-683.

Borenstein, S.(2011): "What Happened to Airline Market Power,"Working Paper.

Brueckner, J. K. (2001): "The Economics of Internaitonal Codesharing: An Anlysis of Airline Alliances," *International Journal of Industrial Organization*, 19, 1475-1998.

Brueckner, J. K. (2003): "International Airfares in the Age of Alliances: The Effects of Codesharing and Antitrust Immunity," *Review of Economics and Statistics*, 85, 1, 105-118.

Brueckner, J. K. and W. T. Whalen (2000): "The Price Effects of International Airline Alliances," *Journal of Law and Economics*, 43, 503-545.

Busse, M. and M. Rysman (2005): "Competition and Price Discrimination in Yellow Pages Advertising,"*RAND Journal of Economics,* 36, 2, 378-390.

Clerides, S. and A. Michis (2006): "Market Concentration and Nonlinear Pricing: Evidence from Detergent Prices in Six Countries," University of Cyprus Working Paper.

Dai, M., Q. Liu and K. Serfes (2010): "Is the Effect of Competition on Price Dispersion Non-Monotonic? Evidence from the U.S. Airline Industry," Drexel University Working Paper.

Dana, Jr., J. D. (1998): "Advance-Purchase Discounts and Price Discrimination in Competitive Markets,"*Journal of Political Economy,* 106, 2, 395-422.

Dana, Jr., J. D. (1999a): "Using Yield Management to Shift Demand When the Peak Time is Unknown,"*RAND Journal of Economics*, 30, 3, 456-474.

Dana, Jr., J. D. (1999b): "Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure,"*RAND Journal of Economics*, 30, 4, 632-660.

Gale, I. and T. Holmes (1993): “Advance-Purchase Discounts and Monopoly Allocation of Capacity,”*American Economic Review*, 83, 1, 135-146.

Gerardi, K. and A. Shapiro (2009): “Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry,”*Journal of Political Economy*, 117, 1, 1-37.

Giaume, S. and S. Guillou (2004): “Price Discrimination and Concentration in European Airline Markets,” *Journal of Air Transport Management*, 10, 5, 305-310.

Hausman, J. A. (1978): “Specification Tests in Econometrics,” *Econometrica*, 46, 6, 1251–1271.

Hernandez, M. and S. Wiggins (2008): “Nonlinear Pricing and Market Concentration in the U.S. Airline Industry,” Texas A&M University Working Paper.

Hernandez, M. (2011): “Nonlinear Pricing and Competition Intensity in a Hotelling-type Model with Discrete Product and Consumer Types,”*Economics Letters*, 110, 3, 174-177.

Holmes, T. (1989): “The Effects of Third-Degree Price Discrimination in Oligopoly,”*American Economic Review*, 79, 1, 244-250.

Katz, M. (1984): “Price Discrimination and Monopolistic Competition,”*Econometrica*, 52, 6, 1453-1471.

McAfee, R., H. Mialon and S. Mialon (2006): “Does large price discrimination imply great market power?,”*Economics Letters*, 92, 3, 360-367.

Orlov, E. (2011): “How Does the Internet Influence Price Dispersion? Evidence from the Airlines Industry,” *The Journal of Industrial Economics*, 59, 1, 21-37.

Stavins, J. (2001): “Price Discrimination in the Airline Market: The Effect of Market Concentration,”*The Review of Economics and Statistics*, 83, 200-202.

Stole, L. (2007): “Price Discrimination and Competition,” In Mark Armstrong and Robert Porter, eds. *Handbook of Industrial Organization*, Vol. 3, pp. 2221-99. San Diego CA: Elsevier Science Publishers.

Tirole, J. (1988): “Price Discrimination,” In J. Tirole, *The Theory of Industrial Organization*, pp. 133-163. Cambridge, Massachusetts and London, England: The MIT Press.

Transportation Group Internatinoal, LC (2002): “Travel Agents Access to Airline Fares,” *A Research Report Prepared for the National Commission to Ensure Consumer Information and Choice in the Airline Industry.*

Wu, D.-M. (1973):“Alternative Tests of Independence between Stochastic Regressors and Disturbances,”*Econometrica*, 41, 733-750.

Table I-1: Summary Statistics[a]

Variable	Mean	StdDev	Minimum	Maximum
Price	1108.9	248.9119843	442	2966
USown	0.9218081	0.268483	0	1
USorigin	0.8923165	0.3099914	0	1
Hub	0.3204085	0.4666502	0	1
AA	0.2710654	0.4445257	0	1
CO	0.055984	0.229899	0	1
DL	0.2040131	0.402993	0	1
NW	0.0709797	0.2568001	0	1
UA	0.3197658	0.4664024	0	1
Chinese_Airlines	0.0616252	0.2404822	0	1
Other_Airlines	0.0165667	0.1276456	0	1
Advance	47.1306769	31.0432369	1	260
Duration	47.5173522	41.7778894	1	327
PEK	0.4765781	0.4994689	0	1
CUyear2005	0.0175664	0.1313738	0	1
CUyear2006	0.0370608	0.1889177	0	1
CUyear2007	0.0631248	0.243196	0	1
CUyear2008	0.1899457	0.392272	0	1
CUyear2009	0.1970151	0.3977581	0	1
CUyear2010	0.1813767	0.3853438	0	1
CUyear2011	0.3139103	0.464097	0	1
UCyear2005	0.0226364	0.1487466	0	1
UCyear2006	0.0379177	0.1910042	0	1
UCyear2007	0.0869751	0.2818087	0	1
UCyear2008	0.1917309	0.3936766	0	1
UCyear2009	0.2075835	0.4055914	0	1
UCyear2010	0.1392459	0.3462153	0	1
UCyear2011	0.3139103	0.464097	0	1
Weekend	0.9367324	0.7474848	0	2
Distance	7143.51	682.6116168	3818	11613.5
Wait	221.8097329	193.988848	0	1032.5
Stops	0.9279849	0.629083	0	3
Slot	0.0309197	0.1731065	0	1
Peak	1.3375464	0.8123791	0	2
Direct	0.2295059	0.4205301	0	1
Total	5.8720366	2.5918628	0	12
Exact	2.2629249	2.1722233	0	10
Better	0.4278777	1.177103	0	11
Worse	4.5297772	2.772306	0	12
Load	0.8370728	0.0819974	0.4805829	1
Domo	0.2685259	0.1443576	0	0.6741351
Population	5691.41	5680.54	7.708	19069.8
Chinese	139.3429874	214.0866643	0	668.493
Income	3.0567623	0.602424	1.2807	7.3929

[a] Income is in $10,000. Population and Chinese are in 1,000s. Number of Observations is 14,004.

Table I-2: Regression Results without Interaction Terms[a]

Variables	OLS Estimated Coefficient	t-stat	IV Estimated Coefficient	t-stat
Adv8_14	-32.4512***	-2.83	-32.9301***	-2.84
Adv15_21	-50.6058***	-4.63	-54.0275***	-4.74
Adv22_28	-51.6758***	-4.86	-50.8575***	-4.71
Adv29_35	-60.5850***	-5.78	-61.1034***	-5.71
Adv36_45	-40.0854***	-3.93	-40.2055***	-3.89
Adv46_60	-50.8637***	-5.02	-51.0475***	-4.95
Adv61_90	-52.6940***	-5.27	-54.9760***	-5.34
Adv91_120	-88.8310***	-8.09	-90.8286***	-8.01
Adv121_150	-84.9832***	-5.20	-74.7233***	-4.12
Adv151_180	-81.0307**	-2.25	-91.8470**	-2.57
Adv181_over	87.8319**	2.36	116.0419**	2.54
Dur6_13	-15.2668	-0.40	-43.1897	-1.01
Dur14_21	-62.1668*	-1.66	-94.5708**	-2.15
Dur22_30	-98.4271***	-2.63	-128.3094***	-2.98
Dur31_45	-79.3311**	-2.11	-110.6916**	-2.54
Dur46_60	-80.3000**	-2.12	-110.7117**	-2.54
Dur61_92	-45.4067	-1.20	-78.4402*	-1.76
Dur93_120	-27.1156	-0.70	-60.2274	-1.34
Dur121_183	-20.2704	-0.53	-57.5467	-1.24
Dur184_over	13.9518	0.31	-56.4341	-0.84
Weekend	16.6341***	7.55	18.0190***	7.32
Wait	0.0244*	1.90	0.0069	0.39
Stops	-20.4814**	-1.98	-18.2063*	-1.71
Direct	3.5545	0.28	15.7575	1.00
Total	-3.6470*	-1.85	-69.3326	-1.56
Distance	-0.0833	-1.40	-0.0247	-0.35
Distance^2	0.0000	1.26	0.0000	0.27
Slot	-23.1352	-1.28	22.6879	0.63
Load	47.5724	1.32	-34.2849	-0.50
Domo	92.8928***	3.80	40.9489	0.96
Income	-11.2767*	-1.65	-18.7905**	-2.17
Chinese	0.0641	0.38	0.7550	1.49
Hub	3.1687	0.38	3.4492	0.40
AA	-52.1370***	-8.49	-22.8202	-1.12
CO	-50.2383***	-4.97	-15.0201	-0.58
DL	-64.9501***	-8.49	-21.8497	-0.74
NW	-40.6518***	-3.11	-1.7175	-0.06
Chinese_airlines	10.0673	0.85	34.8297*	1.68
Other_airlines	-42.4638**	-2.44	-12.2808	-0.46
Constant	966.4770***	3.84	1,263.9734***	3.93
Adjusted R^2	0.605		0.554	
AP χ^2 statistic	16.49			
p-value	0.0000			
AP F test statistic	15.52			

[a] Number of observations is 10,038. Statistical significance at the 1%, 5%, and 10% level are represented by ***, **, and *, respectively. Route and monthly dummies are included.

Table I-3: Averages of Estimated Coefficients for Year-Month Dummy Variables[a]

	OLS		IV	
	peak	off-peak	peak	off-peak
2005	107.29	83.20	-101.15	-172.98
2006	125.53	72.68	-63.89	-129.22
2007	154.42	114.13	-34.94	-57.77
2008	240.37	146.49	83.31	-0.45
2009	172.16	129.72	36.54	-1.76
2010	269.09	136.69	137.25	3.66
2011	304.89	158.51	145.01	12.41

[a] This is a summary of the parameter estimates on monthly dummies. I group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Table I-4: Regression Results with Interaction Terms _OLS[a]

OLS					
	Restriction			Restriction*Total	
Variables	**Estimated Coefficient**	**t-stat**	**Variables**	**Estimated Coefficient**	**t-stat**
Adv8_14	-30.2334	-0.95	Adv8_14*Total	-0.5412	-0.12
Adv15_21	-66.5986**	-2.18	Adv15_21*Total	2.4102	0.55
Adv22_28	-77.3830**	-2.50	Adv22_28*Total	3.9301	0.90
Adv29_35	-62.5834**	-2.07	Adv29_35*Total	0.2278	0.05
Adv36_45	-36.9373	-1.25	Adv36_45*Total	-0.5313	-0.13
Adv46_60	-60.4275**	-2.07	Adv46_60*Total	1.4353	0.35
Adv61_90	-35.2858	-1.24	Adv61_90*Total	-2.9956	-0.74
Adv91_120	-79.7670***	-2.63	Adv91_120*Total	-1.5820	-0.36
Adv121_150	-82.0330*	-1.83	Adv121_150*Total	-0.5526	-0.09
Adv151_180	-95.2001	-1.40	Adv151_180*Total	1.7453	0.15
Adv181_over	272.9259**	2.11	Adv181_over*Total	-25.0006*	-1.65
Dur6_13	-224.7203***	-2.98	Dur6_13*Total	32.2403***	3.09
Dur14_21	-267.2187***	-3.65	Dur14_21*Total	31.6084***	3.11
Dur22_30	-306.5701***	-4.19	Dur22_30*Total	32.1738***	3.17
Dur31_45	-286.2980***	-3.90	Dur31_45*Total	32.0427***	3.14
Dur46_60	-286.7956***	-3.86	Dur46_60*Total	32.0052***	3.10
Dur61_92	-240.2099***	-3.24	Dur61_92*Total	30.1230***	2.94
Dur93_120	-218.4416***	-2.90	Dur93_120*Total	29.4828***	2.83
Dur121_183	-209.7824***	-2.70	Dur121_183*Total	29.3829***	2.74
Dur184_over	-257.2257**	-2.53	Dur184_over*Total	42.2711***	2.97
Weekend	26.0329***	4.65	Weekend*Total	-1.5096*	-1.75
Wait	-0.0219	-0.69	Wait*Total	0.0088	1.60
Stops	-16.9912	-1.02	Stops*Total	0.0877	0.03
Direct	-16.2075	-0.65	Direct*Total	4.5892	1.40
			Peak*Total	2.4236***	2.82
Total	-39.5716***	-3.61			
Distance	-0.0957	-1.61			
Distancesq	0.0000	1.46			
Slot	-25.1437	-1.39			
Load	43.4355	1.22			
Domo	92.6884***	3.81			
Income	-11.3583	-1.64			
Chinese	0.0806	0.46			
Hub	3.4237	0.41			
AA	-51.5632***	-8.41			
CO	-49.3976***	-4.90			
DL	-63.1843***	-8.09			
NW	-42.6915***	-3.49			
Chinese_Airlines	10.4737	0.86			
Other_Airlines	-41.7430**	-2.38			
Adjusted R^2			0.606		

[a] Number of observations is 10,038. Statistical significance at the 1%, 5%, and 10% level are represented by ***, **, and *, respectively. Route and monthly dummies are included.

Table I-5: Regression Results with Interaction Terms_2SLS[a]

2SLS					
	Restriction			Restriction*Total	
Variables	Estimated Coefficient	t-stat	Variables	Estimated Coefficient	t-stat
Adv8_14	-37.6109	-0.88	Adv8_14*Total	0.5044	0.08
Adv15_21	-76.8057*	-1.85	Adv15_21*Total	3.6445	0.59
Adv22_28	-67.0133*	-1.66	Adv22_28*Total	2.4030	0.40
Adv29_35	-54.9783	-1.40	Adv29_35*Total	-1.0595	-0.18
Adv36_45	-35.7876	-0.94	Adv36_45*Total	-0.7394	-0.13
Adv46_60	-64.7840*	-1.69	Adv46_60*Total	2.0940	0.37
Adv61_90	-50.3454	-1.34	Adv61_90*Total	-0.9413	-0.17
Adv91_120	-121.9274***	-2.87	Adv91_120*Total	4.8287	0.76
Adv121_150	-98.3334	-1.60	Adv121_150*Total	3.3544	0.38
Adv151_180	-317.9891***	-2.59	Adv151_180*Total	34.0015**	2.00
Adv181_over	266.2416*	1.73	Adv181_over*Total	-20.6446	-1.20
Dur6_13	-285.4583***	-2.96	Dur6_13*Total	38.1666***	3.12
Dur14_21	-306.4273***	-3.24	Dur14_21*Total	33.5307***	2.85
Dur22_30	-338.5163***	-3.63	Dur22_30*Total	33.2765***	2.84
Dur31_45	-313.9460***	-3.32	Dur31_45*Total	32.1692***	2.72
Dur46_60	-324.3260***	-3.38	Dur46_60*Total	33.9171***	2.83
Dur61_92	-257.5280***	-2.68	Dur61_92*Total	28.6459**	2.40
Dur93_120	-252.7363**	-2.54	Dur93_120*Total	30.5958**	2.48
Dur121_183	-263.6486**	-2.52	Dur121_183*Total	32.9600**	2.51
Dur184_over	-315.9727**	-2.14	Dur184_over*Total	42.2090**	2.32
Weekend	16.9662*	1.88	Weekend*Total	0.1277	0.09
Wait	-0.0239	-0.46	Wait*Total	0.0061	0.62
Stops	-40.1873*	-1.67	Stops*Total	5.3426	1.29
Direct	-47.0452	-0.75	Direct*Total	11.1509	1.38
			Peak*Total	0.5106	0.43
Total	-100.7668**	-2.33			
Distance	-0.0378	-0.55			
Distancesq	0.0000	0.49			
Slot	14.7460	0.42			
Load	-30.4661	-0.48			
Domo	48.3592	1.22			
Income	-17.4704**	-2.12			
Chinese	0.5278	1.10			
Hub	4.5538	0.54			
AA	-27.5311	-1.50			
CO	-20.5932	-0.88			
DL	-24.8955	-0.95			
NW	-8.0441	-0.30			
Chinese_Airlines	28.1096	1.39			
Other_Airlines	-16.0899	-0.64			
Adjusted R^2			0.559		

[a] Number of observations is 10,038. Statistical significance at the 1%, 5%, and 10% level are represented by ***, **, and *, respectively. Route and monthly dummies are included.

Table I-6: Selected Results for first-stage regressions[a]

Variable	AP χ^2 statistic	p-value	AP F test statistic
Total	17.50	0.0000	16.43
Adv8_14*Total	1372.06	0.0000	1288.27
Adv15_21*Total	1228.10	0.0000	1153.10
Adv22_28*Total	1337.04	0.0000	1255.39
Adv29_35*Total	1651.54	0.0000	1550.68
Adv36_45*Total	2125.48	0.0000	1995.68
Adv46_60*Total	2379.46	0.0000	2234.16
Adv61_90*Total	2556.30	0.0000	2400.19
Adv91_120*Total	978.33	0.0000	918.59
Adv121_150*Total	164.53	0.0000	154.48
Adv151_180*Total	40.37	0.0000	37.91
Adv181_over*Total	82.02	0.0000	77.01
Dur6_13*Total	1079.22	0.0000	1013.32
Dur14_21*Total	1637.99	0.0000	1537.96
Dur22_30*Total	1171.69	0.0000	1100.14
Dur31_45*Total	1609.21	0.0000	1510.94
Dur46_60*Total	996.36	0.0000	935.51
Dur61_92*Total	1767.37	0.0000	1659.44
Dur93_120*Total	900.01	0.0000	845.05
Dur121_183*Total	771.45	0.0000	724.34
Dur184_over*Total	127.97	0.0000	120.15
Weekend*Total	3848.48	0.0000	3613.46
Wait*Total	716.19	0.0000	672.45
Stops*Total	929.08	0.0000	872.34
Direct*Total	610.00	0.0000	572.75
Peak*Total	5551.88	0.0000	5212.84

[a] Number of observations is 10,038.

Table I-7: Robustness Check _OLS^a

OLS					
	Restriction			Restriction*Total	
Variables	**Estimated Coefficient**	**t-stat**	**Variables**	**Estimated Coefficient**	**t-stat**
Adv8_14	-7.2421	-0.28	Adv8_14*Total	-3.0506	-0.80
Adv15_21	-49.5508**	-2.00	Adv15_21*Total	0.4174	0.11
Adv22_28	-82.9435***	-3.38	Adv22_28*Total	4.4167	1.23
Adv29_35	-66.0386***	-2.77	Adv29_35*Total	1.1269	0.32
Adv36_45	-37.2069	-1.59	Adv36_45*Total	-0.5830	-0.17
Adv46_60	-52.3358**	-2.28	Adv46_60*Total	0.3917	0.12
Adv61_90	-25.4902	-1.14	Adv61_90*Total	-4.8170	-1.45
Adv91_120	-66.7844***	-2.75	Adv91_120*Total	-4.1171	-1.13
Adv121_150	-80.4256**	-2.24	Adv121_150*Total	-0.8258	-0.16
Adv151_180	-27.1744	-0.39	Adv151_180*Total	-5.6265	-0.51
Adv181_over	18.9986	0.17	Adv181_over*Total	6.1091	0.46
Dur6_13	-178.0077***	-2.60	Dur6_13*Total	24.2583**	2.55
Dur14_21	-224.5547***	-3.37	Dur14_21*Total	23.8283**	2.56
Dur22_30	-260.6345***	-3.92	Dur22_30*Total	24.3584***	2.63
Dur31_45	-257.2721***	-3.85	Dur31_45*Total	25.5555***	2.74
Dur46_60	-270.1207***	-4.01	Dur46_60*Total	26.6936***	2.84
Dur61_92	-210.2765***	-3.13	Dur61_92*Total	23.0041**	2.46
Dur93_120	-190.6854***	-2.80	Dur93_120*Total	23.0984**	2.43
Dur121_183	-183.1473***	-2.62	Dur121_183*Total	23.5002**	2.40
Dur184_over	-210.1520**	-2.29	Dur184_over*Total	35.6160***	2.71
Weekend	27.4338***	5.61	Weekend*Total	-1.5037*	-1.96
Wait	0.0157	0.59	Wait*Total	0.0036	0.79
Stops	3.2249	0.29	Stops*Total	-0.4639	-0.29
Direct	-15.6350	-0.68	Direct*Total	5.5862*	1.87
			Peak*Total	2.4305***	3.21
Total	-32.9246***	-3.33			
Distance	-0.0387	-0.89			
Distancesq	0.0000	0.97			
Slot	-17.2019	-1.13			
Load	76.3777**	2.57			
Domo	52.0104***	2.75			
Income	-10.5526*	-1.75			
Chinese	0.1100	0.70			
Hub	7.8716	1.17			
AA	-46.8182***	-9.05			
CO	-46.3875***	-5.28			
DL	-44.4982***	-7.07			
NW	-38.5536***	-3.94			
Chinese_Airlines	16.2019	1.54			
Other_Airlines	-44.1685***	-3.30			
Adjusted R^2			0.594		

[a] Number of observations is 13,222. Statistical significance at the 1%, 5%, and 10% level are represented by ***, **, and *, respectively. Route and monthly dummies are included.

Table I-8: Robustness Check _2SLS[a]

2SLS					
	Restriction			Restriction*Total	
Variables	Estimated Coefficient	t-stat	Variables	Estimated Coefficient	t-stat
Adv8_14	-9.6143	-0.28	Adv8_14*Total	-2.5389	-0.48
Adv15_21	-38.0690	-1.13	Adv15_21*Total	-1.3517	-0.26
Adv22_28	-77.3945**	-2.39	Adv22_28*Total	3.7015	0.75
Adv29_35	-52.9461*	-1.69	Adv29_35*Total	-0.7084	-0.15
Adv36_45	-27.6186	-0.90	Adv36_45*Total	-2.0487	-0.43
Adv46_60	-50.6783*	-1.65	Adv46_60*Total	0.2932	0.06
Adv61_90	-28.1527	-0.96	Adv61_90*Total	-4.3702	-0.95
Adv91_120	-91.1287***	-2.72	Adv91_120*Total	0.1988	0.04
Adv121_150	-80.8610	-1.62	Adv121_150*Total	0.9324	0.13
Adv151_180	-179.3476	-1.64	Adv151_180*Total	17.7783	1.09
Adv181_over	72.2921	0.58	Adv181_over*Total	0.0262	0.00
Dur6_13	-242.6692***	-2.89	Dur6_13*Total	31.9810***	2.83
Dur14_21	-270.8666***	-3.31	Dur14_21*Total	28.2185***	2.58
Dur22_30	-301.5966***	-3.71	Dur22_30*Total	28.2030***	2.58
Dur31_45	-282.8561***	-3.46	Dur31_45*Total	26.7863**	2.44
Dur46_60	-326.5342***	-3.93	Dur46_60*Total	32.7781***	2.96
Dur61_92	-241.8184***	-2.92	Dur61_92*Total	24.9940**	2.26
Dur93_120	-230.7000***	-2.74	Dur93_120*Total	26.2689**	2.34
Dur121_183	-246.3453***	-2.78	Dur121_183*Total	30.0053**	2.52
Dur184_over	-211.3018*	-1.74	Dur184_over*Total	29.6025*	1.78
Weekend	20.5513***	2.78	Weekend*Total	-0.2731	-0.22
Wait	-0.0012	-0.03	Wait*Total	0.0042	0.56
Stops	-11.3894	-0.74	Stops*Total	2.6718	1.04
Direct	-91.3653	-1.54	Direct*Total	17.0821**	2.22
			Peak*Total	0.5227	0.51
Total	-83.0575***	-3.08			
Distance	-0.0223	-0.51			
Distancesq	0.0000	0.69			
Slot	4.3108	0.21			
Load	17.3500	0.39			
Domo	40.1716*	1.94			
Income	-16.3939**	-2.38			
Chinese	0.3560	1.15			
Hub	6.0686	0.88			
AA	-30.8491***	-2.92			
CO	-29.1506**	-2.09			
DL	-19.6540	-1.48			
NW	-19.2029	-1.32			
Chinese_Airlines	31.4346*	1.93			
Other_Airlines	-20.2643	-1.10			
Adjusted R^2			0.569		

[a] Number of observations is 13,222. Statistical significance at the 1%, 5%, and 10% level are represented by ***, **, and *, respectively. Route and monthly dummies are included.

Table I-9: Robustness Check _Selected Results for first-stage regressions[a]

Variable	AP χ^2 statistic	p-value	AP F test statistic
Total	45.23	0.0000	42.95
Adv8_14*Total	1826.12	0.0000	1734.14
Adv15_21*Total	1585.39	0.0000	1505.53
Adv22_28*Total	1720.87	0.0000	1634.19
Adv29_35*Total	2067.79	0.0000	1963.63
Adv36_45*Total	2459.31	0.0000	2335.44
Adv46_60*Total	3031.47	0.0000	2878.77
Adv61_90*Total	3550.96	0.0000	3372.09
Adv91_120*Total	1108.03	0.0000	1052.22
Adv121_150*Total	172.34	0.0000	163.66
Adv151_180*Total	43.06	0.0000	40.89
Adv181_over*Total	41.82	0.0000	39.71
Dur6_13*Total	1324.57	0.0000	1257.85
Dur14_21*Total	2093.76	0.0000	1988.30
Dur22_30*Total	1292.35	0.0000	1227.25
Dur31_45*Total	1937.49	0.0000	1839.89
Dur46_60*Total	1419.18	0.0000	1347.69
Dur61_92*Total	1881.54	0.0000	1786.77
Dur93_120*Total	1078.20	0.0000	1023.89
Dur121_183*Total	937.52	0.0000	890.30
Dur184_over*Total	130.19	0.0000	123.63
Weekend*Total	5454.10	0.0000	5179.37
Wait*Total	1229.64	0.0000	1167.71
Stops*Total	1460.72	0.0000	1387.14
Direct*Total	689.79	0.0000	655.04
Peak*Total	7092.14	0.0000	6734.90

[a] Number of observations is 13,222.

Chapter 2

Using Search Costs as Another Fence for Market Segmentation: Evidence from the US-China Flight Market

1. Introduction

Price discrimination is a practice firms often use to extract consumer surplus. Airline pricing is a prominent example of this practice. It is a well-known fact that two passengers on the same flight with similar seat positions may have paid very different prices. Airlines price discriminate by using various ticket restrictions to segment consumers into different groups. The restrictions used include whether the flight is direct or not, time-of-day, day-of-week, refundability, advance purchase requirement, cabin and booking class, Saturday-night stayover requirement, etc.

One of the main issues in this literature is the relationship between market structure and the ability or the incentives of firms to price discriminate. The conventional wisdom is that a competitive firm cannot price discriminate because it is a price taker and a firm with market power can price discriminate as long as it can segment the consumers. Hence, the prediction is that as the market becomes more competitive, firms price discriminate less. However, both theoretical studies by Katz (1984), Borenstein (1985) and Dana (1998) and empirical studies by Shepard (1991) and Graddy (1995) show that price discrimination can exist in quite competitive markets. And more recently, Yang and Ye (2008) and Hernandez and Wiggins (2008) show that in second-degree price discrimination models, the relationship between market competition and price discrimination can be positive instead of negative.

Finally, Stole (2007, pp. 2235-2236) compares the oligopoly third-degree price discrimination model by Holmes (1989) with the monopoly third-degree price discrimination model by Robinson (1933) and concludes that the effect of competition on price dispersion across different markets (third-degree price discrimination) is ambiguous and depends on the cross-price elasticities between the products by different firms.

As theoretical studies do not produce a clear prediction, the relationship between market structure and price discrimination is fundamentally an empirical question. Several studies have examined this issue in the airline industry and the results are mixed. Stavins (2001) and Giaume and Guillou (2004) study the price discrimination with respect to the advance purchase requirement and the Saturday-night stayover restriction and find the relationship to be positive. Hernandez and Wiggins (2008) group tickets into 5 quality categories according to their cabin and booking class, refundability and specific travel and/or stay restrictions. They find that price discrimination increases with competition when they compare high-quality tickets to low-quality tickets, but decreases with competition when they compare medium-quality tickets to low-quality tickets. Another strand of the literature examines the relationship between market structure and overall price dispersion, rather than price discrimination with respect to specific ticket restrictions, with the assumption that price discrimination is the main driver of price dispersion.[40] Again, the results are mixed. Borenstein and Rose (1994) find that more competition leads to more dispersion, while Gerardi and Shapiro (2009) and Dai, Liu and Serfes (2012) find the relationship to be

---

[40] Other explanations for price dispersion include peak-load pricing (e.g. Dana 1999a) and demand uncertainty and costly capacity (e.g. Dana 1999b).

negative and nonlinear, respectively.[41]

Our paper offers new evidence on the relationship between price discrimination and market structure using data from a new market. Unlike studies cited above that focus on how airlines use ticket restrictions to price discriminate in the more mature US and European markets, we focus on how airlines exploit imperfect information and consumers' search costs as a fence to segment consumers and price discriminate in the emerging US-China flight market.

In the US-China flight market, airlines contract with consolidators (wholesale travel agents) to sell a large number of tickets for them. In return for their service, they offer consolidators a large discount on ticket price. Consolidators then work with retail travel agents to market tickets to travelers and the discount offered by the airline is usually shared among all three parties. As a result, consumers are segmented into two groups based on their search costs. One group of consumers, who have low search costs, search for the low price and purchase the ticket through a travel agent. The other group of consumers, who have high search costs, do not search for the low price and purchase the ticket at a higher price from another source like the airline's official website. Theoretical studies of this practice can be dated back to Salop (1977), who shows that a monopolist can open multiple sales outlets and offer different prices at each in order to price discriminate against consumers with high search costs. Though third-degree price discrimination models by Borenstein (1985) and

---

[41] Studies of other industries also find mixed results regarding the relationship between price discrimination and market structure. Busse and Rysman (2005) find that more competition leads to less price discrimination in the yellow pages advertising market. In contrast, Asplund, Eriksson and Strand (2008) and Borzekowski, Thomadsen and Taragin (2009) find the relationship to be positive in the newspaper industry and the market for mailing lists, respectively. Finally, Clerides and Michis (2006) study the detergent market in six countries and find the relationship to be positive in some countries, while negative in others.

Holmes (1989) are related, Stole (2007, footnote 7) points out that competitive analogs of Salop's monopoly model, where consumers' search costs play an explicit role, have not been well explored. Also, there is a large theoretical (e.g. Varian 1980; Rosenthal 1980; Stahl 1989) as well as empirical (e.g. Baye and Morgan 2004; Hortacsu and Syverson 2004; Hong and Shum 2006) literature that uses search costs to explain inter-firm rather than intra-firm price dispersion. However, to the best of our knowledge, no empirical study has examined firms using search costs as a fence to price discriminate.

Using a unique dataset collected from the US-China flight market, we first quantify the magnitude of the price difference between different ticket distribution channels. We find that for the same ticket, the difference between the agent price and the price from the airline's official website is about 46% of the agent price on average. We then further examine the relationship between price discrimination and market competition using regression analysis. Under mild identification assumptions, we find evidence that price discrimination with respect to costly consumer search increases with market competition.

The rest of the paper is organized as follows. Section 2 discusses the unique aspects of the international flight market between the US and China. Section 3 introduces the data. Our empirical models and identification strategies are detailed in Section 4. Results are reported and discussed in Section 5. Finally, Section 6 concludes.

2. The International Flight Market between the US and China

The US-China international flight market has expanded significantly during the last three decades due in large part to China's Reform and Opening-up policy of late 1978. The Civil Aviation Administration of China (CAAC) and Pan American World Airways (Pan Am) began flying between mainland China and the US in the early 1980's,[42] and Northwest Airlines and United Airlines joined the trans-Pacific flight market shortly thereafter.[43] However, since Chinese carriers were not competitive in the international flight market at that time, the Chinese authority hesitated to make the market more open to foreign airlines. As a result, the price of a US-China flight ticket was relatively high and demand was low during the 1980's.[44] As time passed, the rapid economic growth of China stimulated huge demand for international flight services between the US and China. Northwest Airlines began service of the first scheduled non-stop flight between the two countries in 1996.[45] More substantial progress of the passenger air service was made in the 1999 Air Services Agreement, which asserted that four airlines from each country were allowed to offer flight services between the US and China. This doubled the weekly number of flights from each country from 27 to 54.[46] Further expansion of the US-China flight market was made in a new agreement in July 2004, which relaxed the restriction that flight services were only allowed between specific cities. Unlimited code-sharing between US and Chinese air carriers was permitted.[47] Another bilateral agreement was reached in July 2007, with the main achievement that by end of 2011, the number of airlines in each country allowed to serve the trans-Pacific market would increase from four to nine and the number of weekly

---

[42] Source: http://query.nytimes.com/gst/fullpage.html?res=9903E0DC123BF93AA15752C0A967948260. Visiting date: 08/08/2012.

[43] Source: http://www.everythingpanam.com/. Visiting date: 08/08/2012.

[44] Source: http://www.voanews.com/tibetan-english/news/a-28-a-2004-07-24-2-1-90264472.html.

[45] Source: Northwest Historical Timeline (1990) from the webpage of Delta Airline. Visiting date: 08/08/2012.

[46] Source: http://transportation.northwestern.edu/. Visiting date: 08/08/2012.

[47] Source: http://goliath.ecnext.com/coms2/gi_0199-555172/United-States-China-sign-agreement.html. Visiting date: 08/08/2012.

flights would increase from 54 to 249 from each country.[48] Due to the unfavorable market condition in international travel industry caused by the subprime mortgage crisis in the US, the new cap has not been reached as of today. But several new entrants including Delta, US Airways and Hainan Airline (one of the Chinese airlines) started their services in the second half of 2010.

In recent years, there have been numerous studies examining various aspects of the US domestic flight market. Examples include Borenstein (1989), Borenstein and Rose (1994), Dana (1998, 1999a, 1999b) and Gerardi and Shapiro (2009), to name just a few. There has been far less research on international flight markets. Some examples include Brueckner (2001, 2003) and Brueckner and Whalen (2000). Although the international passenger air market between the US and China shares quite a few of characteristics as the US domestic market, it is unique in several aspects, with the biggest difference in how tickets are sold.

Currently, airlines sell their tickets through their own Computer Reservation Systems (CRS's). CRS is a computerized system used to store and retrieve marketing information and conduct transactions related to air travel.[49] It interfaces with airline's inventory system and supports the operational functionality of airlines by adjusting fares based on changes in inventory and current market conditions. When a traveler calls ticketing office or visits website of an airline and purchases a ticket there, the sale is completed using the airline's own CRS directly. The airline does not need to pay any commissions or booking fees to third-party business partners.

---

[48] Source: http://www.uschinatrip.com/news.cfm. Visiting date: 08/08/2012.

[49] Source: http://en.wikipedia.org/wiki/Computer_reservations_system. Visiting date: 08/08/2012.

Airlines also sell tickets by subscribing to and linking their CRS's, or part of their CRS's to one or several Global Distribution Systems (GDS's) and selling tickets with the help of travel agents. A GDS is maintained by a third party data distributor and is essentially a platform that collects ticket information such as fare, schedule and availability of different seats from providers and displays the results globally to travel agents who subscribe to it.[50] Travel agents who subscribe to GDS's are able to offer their customers a variety of airfares and services and are able to confirm ticket reservations with airline's CRS in real time. GDS's are therefore great platforms for airlines to sell their tickets as well as to monitor the pricing of similar tickets by their competitors. Selling tickets through GDS's generates costs to airlines. The airlines need to pay a subscription fee to the GDS's. In addition, every time a ticket is sold through a GDS, the airline pays a booking fee to the GDS (Orlov 2011).

Before the arrival of the internet age, selling tickets through GDS's with the help of brick and mortar travel agencies was the dominant channel for airlines to distribute their tickets, as that was the easiest way for consumers to search for and compare prices from different airlines. Orlov (2011) reports that approximately 80 percent of all airline tickets were sold through brick and mortar travel agencies by the mid-1990's. Selling tickets through travel agents was not free and airlines had to pay commissions for tickets sold. With the arrival of the internet age, airlines developed their own websites and internet-based third-party businesses also emerged (Transportation Group International, 2002).[51] This made it

---

[50] GDS's also include information on hotel, car rental and other perishable services. Examples of GDS's include Sabre in the US, Apollo and Amadeus in Europe, Abacus in Japan and TravelSky in China.

[51] These include online travel agencies (e.g. expedia.com), price search engines (e.g. farechase.com) and bid-based websites (e.g. priceline.com). Different internet-based businesses have different sources of ticket information, depending on their business models. Some subscribe to GDS's, some have access to airlines'

relatively easy for travelers to search for and compare prices by themselves. At the same time, airlines gradually phased out commissions paid to the travel agents for selling tickets for domestic travel and the travel agents had to start charging travelers fees for their services. As a result, the demand for services by travel agencies in the domestic flight market has decreased significantly and currently, consumers rarely use these agencies for the sole purpose of purchasing a domestic flight ticket, though some consumers still use them when they purchase a non-traditional ticket such as an open-jaw ticket, or arrange for a vacation for which a ticket is often bundled with other products like hotel stays and car rentals.

For the international flight market between the US and China, however, airlines keep offering travel agents commissions and travel agencies still play a major role in distributing tickets, despite the rapid development of internet, for the following reasons. First and foremost, working with travel agents provides airlines a way to get around the fare regulations imposed by the International Air Transport Association (IATA) and practice third-degree price discrimination.[52] IATA is an international industry trade group of airlines formed in 1945 with the mission to represent, lead and serve the airline industry. Currently, it has more than 240 members from more than 110 countries. Its main activity is to serve as the price setting body for international airfare. In an arrangement going back to 1944, international airfare prices have been set through bilateral governmental agreements rather than through market mechanisms. Airlines had been granted a special exemption by major antitrust agencies in the world to consult prices with each other through this body. As a

---

CRS's through special arrangement and some simply search dozens of websites including the websites of airlines as well as the websites of other online travel agencies.

[52] The rest of this paragraph largely follows the content at http://en.wikipedia.org/wiki/International_Air_Transport_Association. Visiting date: 08/08/2012.

result, originally both domestic and international airfares were highly regulated by IATA. Since the deregulation in the US in 1978 and later in Europe, the US and many European countries signed bilateral "open skies" agreements that weakened IATA's price fixing role. At the same time, antitrust investigations were launched both in the US and Europe. Under these pressures, IATA withdrew fare regulations within EU and between EU and the rest of the world in 2006 and 2007, respectively. However, the international airfares for other markets, including the US-China flight market, remains regulated by IATA.[53]

Under IATA fare regulations, airlines are not allowed to sell tickets at prices below the minimum price set by IATA.[54] This is costly to the airlines as they cannot deeply discount the tickets when necessary to fill the capacity. This is especially true in the US-China flight market as a trans-pacific flight is usually served by a big airplane with at least 250 seats, which consumes a large amount of fuel. But interestingly, IATA does not regulate the amount of commission an airline can pay the travel agents. Because of this loophole, airlines can get around of the IATA fare regulations and charge lower prices by offering travel agents a large amount of commission, part of which are then passed onto the consumers. To achieve this, for each market, each airline usually works with a few consolidators. Consolidators are a special type of travel agents who specialize in buying tickets in bulk from airlines at a deeply discounted price and they usually do not work directly with consumers.[55] Consolidators then resell the tickets to retail travel agents, who

---

[53] For a complete list of the countries that participate in the US "open skies" program, see www.state.gov/e/eb/rls/othr/ata/114805.htm. Visiting date: 08/08/2012.

[54] IATA does not regulate the maximum price.

[55] Consolidators play an important role in international flight markets, but only a minor role in the US domestic flight market. To learn more about consolidators, see the article at

then resell the tickets to consumers. As a result, the large discount offered by an airline for a ticket is shared by the consolidator, the travel agent and the consumer. The relationships among airlines, consolidators, travel agents, and travelers are essentially the same as those of producers, wholesalers, retailers and consumers in other markets. Practically, airlines sell these deeply discounted tickets through their own CRS's. Based on the contracts or agreements between the airlines and the consolidators, during a specific time period (usually two or three months before the departure date during peak season and one month before the departure time during off-peak season), airlines make some tickets with large commissions available to the consolidators through an expanded menu of options in their CRS's. The consolidators then pass the access rights to these heavily discounted tickets in CRS's to those retail travel agents who work with them. These retail travel agents then sell the tickets to consumers.[56]

Therefore, in the US-China international flight market, airlines essentially offer tickets at two different prices. The first price is the published fare, which is available through CRS's and GDS's to the general public via official websites of airlines, airport ticket offices and travel agents who do not work with consolidators. The other price is the discounted price, which is only available through travel agents who work with consolidators and the price is usually lower than the published fare.[57] This can be regarded as a practice of third-

---

http://tourism.about.com/od/travelagentsandagencies/a/Travel-Agents-Utilizing-Airfare-Consolidators.htm. Visiting date: 08/08/2012.

[56] For more information about how this process works, see the article at http://www.smartertravel.com/travel-advice/the-truth-about-consolidator-fares.html?id=9581452. Visiting date: 08/08/2012.

[57] It is worth noting that airlines can adjust the published fare anytime based on the inventory and current market conditions. Since the discounted price is simply the difference between the published fare and part of the commissions airlines offer consolidators, the discounted price also changes when the published fare changes.

degree price discrimination as with this arrangement, airlines segment the travelers into two groups. One group of travelers, who have low search costs, search for low price and become informed of the fact that cheaper tickets are available through some travel agencies. The other group of travelers, who have high search costs, are uninformed and hence purchase the more expensive tickets from other sources.

This strategy is particularly profitable for the airlines during peak travel seasons. During peak travel seasons, demand for international travel between the US and China by both informed and uninformed travelers is high. Airlines can then offer informed travelers a relatively low price through travel agents and offer uninformed travelers a relatively high price. As a result, a large number of informed travelers and a small number of uninformed travelers with low demand elasticity (leisure travelers with high income or business travelers) purchase the tickets. The airlines fulfill the capacity and earn a large mark-up from those tickets sold to uninformed travelers. During off-peak seasons, demand by both groups of travelers is low. The peak season strategy, if used, would not generate enough demand to fulfill the capacity. As a result, airlines have to offer a relatively low price to both informed and uninformed travelers. Indeed, our results below show that airlines price discriminate more during peak travel seasons.

Another reason that travel agents survive in this market is the fact that a significant percentage of the international flight passengers between the US and China are people with Chinese backgrounds, their relatives, and visitors. Most of them speak Chinese fluently and some have difficulty speaking English and searching websites in English. Therefore, travel agents who speak Chinese can reach and serve these consumers better. Indeed, most of the travel agencies serving this market are run by Chinese Americans who speak Chinese fluently and are located in large metropolitan Chinatowns where many Chinese or Chinese Americans live and visit regularly.

3. Data

The data used in our study come from several different sources. Each data source in explained in turn below with a focus on the variables used in our regression analysis.

3.1 Ticket Data

As discussed in section 2 above, in the US-China flight market, airlines sell tickets at two different prices, one at the published fare amount and the other with a discount. We collected data on both prices for (almost) the same ticket. The data on most of the ticket characteristics and the discounted price were collected from the website of TravelSuperlink.com.[58] This website started in 2005 as a simple platform that allows air travelers between the US and China to share with one another their recent purchasing experiences, with the goal to help travelers find good deals. To contribute to this effort, travelers can upload the information of the ticket they recently purchased through a ready-to-fill form on the website. The information collected includes the purchase date, the departure date, the return date (if it is a round trip), the trip type (round/single/open/open-jaw),[59] the departure city, the arrival city, the connecting cities if there are any, the airfare amount, the airline carrier, the name of the travel agency through which the ticket is booked, and any

---

[58] The specific webpage we collected our data is http://gochina.travelsuperlink.com.

[59] Both open and open-jaw tickets are for round trips. An open ticket is a ticket with a flexible return date. A ticket is open-jaw if the traveler's final destination of the round trip is different from the departure city. For example, the round trip ORD-PEK, PEK-JFK is open-jaw. It can also be the case that the traveler returns from one city other than the one he arrived. For example, the round trip ORD-PEK, PVG-ORD is open-jaw too.

comments regarding the quality of the service provided by the agent. Over the years, this simple platform has gained popularity among the Chinese community and evolved into a full-blown website. It is now not only a place where travelers share their information, but also a place where ticket agents and other businesses in the travel industry advertise their services and special deals. The information a user can get from the website is no longer confined to the US-China air ticket deals, but also includes travel information with destinations to almost all parts of the world and through other transportation modes like cruises and car rentals.

During our data collection period, we visited the website of TravelSuperlink.com every day at the same time. For each new post of flight ticket information, we used the price search engine website FlyChina.com to search for the non-refundable economy-class airfare of the itinerary with the same departure-connection(s)-arrival airport combination, the same departure-return flight dates, the same airline, as well as the same trip type (only roundtrip flights are included in the analysis). FlyChina.com is a price search engine website that visits the official websites of all airlines serving the US-China flight market, collects the price and itinerary information for tickets from all airlines and displays them to the traveler on one screen. For a given ticket, the information from Flychina.com is the same as the one from the official website of the airline and the airfare is the published fare amount. We focused on non-refundable economy-class airfare from FlyChina.com because tickets reported on TravelSuperlink.com were of the same type. For some tickets, there were multiple matches from FlyChina.com. This is mainly due to the fact that airlines often serve multiple connecting flights between the origin/destination airport and the connecting airport(s). The prices for these multiple matches were usually the same or differ very little and we chose the price from the match with the cheapest price.

We focus our analysis on 4,555 round-trip tickets from 294 routes with both departure and return dates between January 1st, 2011 and December 31st, 2011 and with Beijing or Shanghai as the Chinese origin/destination city.[60] A route is defined as a triplet of origin-destination-origin airports. Therefore, PEK-ORD-PEK and ORD-PEK-ORD are two different routes. Using the reported and searched information we created the following variables for each ticket: the discounted airfare amount from TravelSuperlink.com (Price1), the published airfare amount from Flychina.com (Price2), whether the trip is direct (non-stop) or not (Direct), whether the origin or the destination city is one of the hubs of the air carrier (Hub), dummies for air carriers, number of days between the purchase date and the departure date (Advance), number of days between the departure date and the return date (Duration), monthly dummies for the US to China leg of the trip, monthly dummies for the China to US leg of the trip, half of the total flight distance of the round trip (Distance), number of minutes spent at the airport during travel (Wait), number of stops (Stops) and the number of flight dates that are during the weekend (Weekend). Weekend is defined to be Friday, Saturday, Sunday and Monday, not just Saturday and Sunday. Several travel agents told us this is how weekend is defined in this market.[61] Peak travel season is defined to be January, May, June, July, August and December, which correspond to winter and summer vacation times. Our

---

[60] Including tickets to or from other cities in China in the analysis requires data on flight schedules in China, to which we do not have access.

[61] We conjecture this is because connecting flights (flights within the US) on these four days have higher load factor. Connecting flights on Mondays and Fridays have higher load factor because more domestic travelers travel on these two days. Connecting flights on Saturdays and Sundays also have higher load factor because fewer connecting flights are scheduled on Saturdays and Sundays.

summary statistics results below also confirm these months constitute the peak travel season. Finally, we created a variable (Lag) to control for the time difference in days between the ticket purchase date and the date we searched for the published fare. Most of the travelers posted their ticket information on TravelSuperlink.com right after their purchase and we were able to search for the official airfare within 24 hours, but some travelers posted the ticket information some days later and the time difference between the purchase date and the date we searched for the published fare is larger. This resulted in a difference in the Advance variable (number of days between the purchase date and the departure date) between the ticket the traveler reported on TravelSuperlink.com and the ticket we searched on FlyChina.com and that is the reason we refer to our two prices as prices for almost the same ticket, rather than the exactly same ticket.

3.2 Official Airline Guide (OAG) Data

The second data source for our empirical analysis is the worldwide issues of the Official Airline Guide (OAG), which report schedule information for almost all commercial flights in the entire world. For each ticket, using the reported information on departure date, return date, origin and destination airport, we first searched the OAG databases to identify all the feasible itineraries by all airlines that correspond to the ticket. A feasible itinerary is defined as an itinerary that has the same departure and return dates, the same origin and destination airports, a total travel time less than 30 hours, up to 2 stops in the US, up to 1 stop in China and up to 1 stop in a third country like Canada, Japan and Korea, a layover time more than 60 minutes in the connecting airport where passengers need to go through border control and a layover time more than 40 minutes in the connecting airport(s) where passengers do not need to go through border control.

Using the feasible schedules data generated from OAG, we created the variable measuring market competition. The market competition variable is the key explanatory variable in our regression analysis. Previous studies used market shares on a route, either in terms of the number of flights operated or in terms of the number of passengers transported, as measures for market competition. For example, Hernandez and Wiggins (2008) and Gerardi and Shapiro (2009) used the Herfindahl-Hirschman index (HHI) in terms of the number of passengers transported and Stavins (2001) used HHI in terms of the number of flights operated. We cannot use HHI in terms of the number of passengers transported because we do not have data on the numbers of passengers transported by each airline for each route. For example, United-Continental serves the route RDU-PEK-RDU with a connection at EWR. Though data on the number of passengers transported by United-Continental between RDU and EWR is available from Bureau of Transportation Statistics (BTS), we do not know how many of them are travelers heading to PEK. Therefore, we use the number of other airlines (N) that can serve the same route on the same departure and return dates as the measure for market competition. This variable should be highly correlated with a market competition variable based on the number of flights operated, like the one used by Stavins (2001). This is because for most of the routes, each airline only has one daily scheduled flight for the international segment to serve consumers.[62]

3.3 Bureau of Transportation Statistics (BTS) Data

Our third data source is several datasets from the BTS. First, the Data Bank 28DS

---

[62] This is different from the case of trans-Atlantic market. For example, both American Airlines and United Airlines offer several daily flights between New York and London.

and 28IS datasets provide information on the number of passengers transported and the number seats available on non-stop flights between a pair of airports by airline and month. Data Bank 28DS covers airport pairs where both airports are located in the US, while Data Bank 28IS covers airport pairs where one airport is located in the US and the other airport is located outside of the US. For each observation in our ticket data, we first used the BTS data to compute the load factor for each segment of the itinerary. The load factor is defined as the ratio between the total number of passengers transported and the total number of seats available from all flights by the corresponding airline in the corresponding month. The flights included both the flights operated by the airline itself and flights operated by its code share partners. We then computed the weighted average load factor variable (Load) for the ticket by averaging across all the load factors for its segments, using the segment distance as the weight. We were able to create the Load variable for 4,396 out of the 4,555 ticket observations. For 143 tickets, the Load variable could not be computed because the itinerary involves a stop in a third country and as a result, the load factor for the segment between China and the third country could not be computed because the BTS datasets do not cover such flights. For the other 16 tickets, the Load variable was not created simply because no corresponding statistics could be found in the BTS data for some or all of the segments in the itinerary.

Then, we used the Origin and Destination Survey-DB1B dataset to create the airport dominance (Domo) variable. The DB1B dataset provides quarterly information on the number of passenger originations by each domestic airline at each airport in the US.[63] For each observation in our ticket data, we first used the information from the DB1B to compute the airline's shares of passenger originations for all the domestic airports on the itinerary during the corresponding quarter. Then, following Borenstein and Rose (1994) and Borenstein (2011), we took the average of shares across all the airports as the Domo variable. For 190 out of the 4,396 tickets that have non-missing values for all other variables, the airline is a foreign airline and does not report statistics to BTS. We set the Domo variable for these tickets to be zero as foreign airlines (especially those airlines which serve the US-China market) usually account for a small share of passenger originations at US airports.

3.4 Census Data

Finally, we obtained the population and per capita income data from the Census. For each ticket, we got the number of total population (Population), the number of Chinese-ethnicity (except Taiwan origin) population (Chinese) and per capita income (Income) from the corresponding metropolitan and micropolitan statistical areas the US origin/destination airport on the itinerary serves.[64, 65] For each observation, all these data are from the year when the ticket was purchased (either 2010 or 2011).

---

[63] An origination is the beginning of a directional trip.

[64] Some airports serve multiple MSAs. For example, airport MBS, serves Midland, MI, Micro Area, Bay City, MI Metro Area and Saginaw-Saginaw Township North, MI Metro Area.

[65] Though each ticket has two ending cities, one in China and one in the US, we did not average the population and income variables across the two ending cities. This is because Beijing and Shanghai have similar population and per capita income. Therefore, taking averages across the two ending cities would yield variables that have similar cross-observation variations as these variables for the US city alone.

3.5 Summary Statistics

Since our identification assumption (discussed below in Section 4) is more likely to hold for tickets with a small time difference between the ticket purchase date and the date we searched for the published fare, we focus on the 3,892 tickets for which the Lag variable is less than or equal to 2 in our regression analysis. Table 1 provides the summary statistics for the variables that are used in the estimation. Several features of the data are worth mentioning. First, there is significant variation in the observed prices. Price1 (the price from travel agents) ranges from $628 to $2,966, with an average of $1,212. Hence, the range is about 193% of the mean. Part of the variation comes from the fact that the data include both tickets for flights between the west coast and China and between the east coast and China, with flights between the west coast and China cost much less. One reason for this is that the west coast is closer to China and hence a flight between the west coast and China incurs less fuel cost for the airline. A similar pattern is observed in Price2, the published airfare. There, the range is about 670% of the mean. The maximum observed published fare in the dataset is $12,520, for a non-refundable economy class ticket that was bought only 6 days in advance through a travel agent with a price at $1,575.

Second, as expected, the average published fare is about $568 higher than that of the agent fare. This is equivalent to 47% of the average agent fare. Indeed, purchasing tickets through agents can save travelers a significant amount of money. However, it is also worth noting that for 420 tickets, the published fare is actually lower than the agent fare. For these tickets, the published fare is $78 less than the agent fare on average. As discussed in Section 2, airlines only sell some, not all, tickets through consolidators. As a result, sometimes, especially during off-peak season when consolidators contract for a smaller number of tickets in bulk, travel agents run out of deeply discounted tickets. When that happens, agents can only sell consumers tickets at the published fare amount plus a service fee.[66] In these cases, travelers will actually pay a lower price if they purchase from the airline directly. As this happens only occasionally, some travelers may not be aware of this and hence pay a higher price.

Third, although about 96% of the tickets in our sample are sold by the three US airlines AA, Delta and UA, on average, an airline faces 6 competitors on a route. Some routes are served by one single airline, while other routes are fairly competitive, with the maximum number of competitors being 11.

Fourth, for 3,017 of the 3,892 tickets, at least one leg of the trip was during the months of January, May, June, July, August and December, confirming these months constitute the peak travel seasons.

Other data features worth discussing are the following. The Lag variable has an average of 0.9, indicating the time difference between the ticket purchase date and the data we searched for the published fare is within 24 hours for most of the tickets in the sample, making our identification assumption more likely to hold. Airlines from China, Japan, Korea and Canada, only have a combined market share less than 5%. It is difficult for foreign airlines to compete with US airlines because they do not serve many of the medium and small size cities in the US, even though they have code share partners. An average traveler purchases the ticket 45 days in advance with duration of stay about 51 days. This indicates

---

[66] Another reason could be consumers purchase the tickets from agents who do not work with consolidators.

that the majority of the ticket purchasers in this sample are students, retirees and homemakers, rather than full time employees. The layover time for an average traveler is about 3.8 hours and about one third of the tickets are direct flight tickets. Finally, the load factor is 0.86 on average. This is largely due to the fact that the trans-pacific flights are usually fully packed, especially during the peak travel seasons.

4. Empirical Analysis

Our goal here is to study the relationship between price discrimination and market competition. Following Tirole (1988), we define price discrimination as the difference in price-cost margins (i.e. the difference between price and marginal cost). However, in our data, we only observe the difference in prices, not price-cost margins, and hence our regression results only tell us the relationship between price difference and market competition. Therefore, we first need to study under what (identification) assumptions we can infer the relationship between price discrimination and market competition from the relationship between price difference and market competition.

4.1 Identification

For the same ticket, denote p_1 as the price the traveler pays the travel agent, p_2 as the published fare and p_0 as the price the consolidator pays the airline under their agreement. We have the following relationship among the prices,

$$(1).\qquad p_2 - p_1 = p_2 - c_2 - (p_0 - c_0) - (p_1 - p_0) + (c_2 - c_0),$$

where c_0 is the airline's marginal cost of selling and honoring the ticket through the consolidator and c_2 is the airline's marginal cost of selling and honoring the ticket through the other channel. Therefore, $p_2 - c_2 - (p_0 - c_0)$ is the difference in price-cost margins for selling the same ticket through the two channels and hence is our measure of third-degree price discrimination in this study. Finally, $M_1 = p_1 - p_0$ in (1) is the total commission earned by the consolidator and the travel agent if the ticket is sold through them.

As Dana (1998) points out, the marginal cost of honoring a ticket in the airline industry is better described as the sum of the marginal cost of production, which is incurred only when the traveler is on board like the cost of serving meals and drinks, and a shadow cost of capacity, which is incurred whether or not the ticket is actually sold. The shadow cost is higher when more seats of a flight are sold because the opportunity cost of each remaining seat is higher. Hence, the same seat or ticket has different shadow costs of capacity at different times. As we collected the two prices for the same ticket at almost the same time, the marginal costs of honoring the ticket are likely to be same for the two ticket distribution channels.[67] As a result, (1) becomes,

$$\text{(2).} \quad M_2 = p_2 - (pc + sc_2) - [p_0 - (pc + sc_0)] - (p_1 - p_0) + (pc + sc_2) - (pc + sc_0)$$

$$= PD - M_1 + (sc_2 - sc_0),$$

where $M_2 = p_2 - p_1$, pc is the marginal cost of production, sc_0 and sc_2 are the marginal selling costs of the same ticket through the two distribution channels, and

$PD = p_2 - sc_2 - (p_0 - sc_0)$.

Differentiating both sides of (2) with respect to N, the number of competitors on the route, yields,

---

[67] As discussed above, the small difference between the ticket purchase date and the date we searched for the published fare is controlled using the Lag variable in regression analysis.

(3). $$\frac{\partial M_2}{\partial N}=\frac{\partial PD}{\partial N}-\frac{\partial\left[M_1-\left(sc_2-sc_0\right)\right]}{\partial N}.$$

We are interested in $\operatorname{sgn}\left(\frac{\partial PD}{\partial N}\right)$, where $\operatorname{sgn}(\cdot)$ is the sign function. However, we are only able to obtain $\operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right)$ from our regression analysis. If $\operatorname{sgn}\left(\frac{\partial PD}{\partial N}\right)=\operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right)$, then we can infer the relationship between third-degree price discrimination and market competition from the estimated relationship between price difference and market competition. As $PD=M-\left(sc_2-sc_0\right)$, where $M=M_1+M_2$, a sufficient condition for this is,

(4). $$\operatorname{sgn}\left(\frac{\partial\left(sc_2-sc_0\right)}{\partial N}\right)=0 \text{ and } \operatorname{sgn}\left(\frac{\partial M}{\partial N}\right)=\operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right).$$

This is because when $\operatorname{sgn}\left(\frac{\partial\left(sc_2-sc_0\right)}{\partial N}\right)=0$, $\operatorname{sgn}\left(\frac{\partial M}{\partial N}\right)=\operatorname{sgn}(PD)$. And if $\operatorname{sgn}\left(\frac{\partial M}{\partial N}\right)=\operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right)$, then we have $\operatorname{sgn}\left(\frac{\partial PD}{\partial N}\right)=\operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right)$.

In our particular context, (4) is likely to hold. We first discuss the first part of (4). sc_0 is the airline's marginal cost of selling a ticket through consolidators. Note sc_0 does not involve the commission the airline pays the consolidators as p_0 is defined as the price the airline receives from the consolidators. To sell tickets through consolidators, an airline needs to incur costs to negotiate and sign agreements with the consolidators and also maintains its CRS so that consolidators can access its CRS to make ticket reservations. Both costs are unlikely to vary across different routes. First, consolidators and airlines do not negotiate and sign contracts route by route. Usually, they sign one contract that covers all the routes between the US and China. Second, CRS is a network wide system and hence its maintenance cost is essentially an overhead cost for all routes. sc_2 is the marginal cost of selling a ticket at the published fare. To sell tickets through this channel, the airline needs to maintain its CRS. Also, it needs to pay a subscription fee to one or several GDS's and pay booking fees to GDS's if the tickets are sold through GDS's. Again, all these costs are unlikely to change by route. Therefore, both sc_0 and sc_2 are likely to be independent of the number of competitors, N, and as a result, $\operatorname{sgn}\left(\frac{\partial(sc_2 - sc_0)}{\partial N}\right) = 0$.

We now turn to the second part of (4), that is, $\operatorname{sgn}\left(\frac{\partial M}{\partial N}\right) = \operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right)$. As discussed in section 2 above, when the airline sells a ticket through the consolidator, it offers the consolidator a large commission. The consolidator then passes part of the commission to the travel agent, who then passes part of the commission to the traveler. Therefore, the commission offered by the airline is shared by the consolidator, the travel agent and the traveler. $M_1 = p_1 - p_0$ is the sum of the consolidator's and the travel agent's shares of the commission, while $M_2 = p_2 - p_1$ is the traveler's share of the commission. As a result, M, is the commission offered by the airline. If the airline increases (decreases) M when the route becomes more (less) competitive, $\frac{\partial M_2}{\partial N}$ is positive (negative) if the pass through rate of the increase (decrease) in commission from the consolidator and the agent to the traveler is strictly larger than 0. In this case, we have $\operatorname{sgn}\left(\frac{\partial M}{\partial N}\right) = \operatorname{sgn}\left(\frac{\partial M_2}{\partial N}\right)$. A pass-through rate of 0 indicates that the consolidator and the agent together enjoy the monopoly market power.

However, this is unlikely in the US-China international flight market. As described above in section 2, in this market, each airline usually works with a few rather than just one consolidators and there are a large number of travel agents. So neither the consolidators nor the travel agents are likely to enjoy monopoly power.

At the other extreme, a pass-through rate of 1 indicates the consolidator and the travel agent do not have the ability to change their price-cost margins. This is actually not an impossible case in the US-China international flight market. In this market, each airline usually works with a few consolidators. Therefore, given the same ticket, consolidators compete with one another in an oligopoly market. As this is a homogenous good market, if the consolidators are identical, then the equilibrium is the Betrand-Nash equilibrium where the equilibrium price equals the marginal cost. Therefore, in equilibrium, each consolidator charges each retail travel agent $p_c = p_0 + c_c$, where p_0 is defined above and c_c is the consolidator's per ticket operating cost. As the consolidators sell tickets for the airline on all routes, it is unlikely for the operating cost to depend on which route the ticket is for and hence the number of competitors on that route. Similarly, in this market, there are a large number of retail travel agents and they compete with one another in a perfectly competitive market to sell the same ticket. As a result, if the retail travel agents are identical, then in equilibrium, each retail travel agent charges each traveler $p_1 = p_c + c_a$ where c_a is the retail travel agent's per ticket operating cost. Again, as the retail travel agents sell tickets for all routes, it is unlikely for the operating cost to depend on which route the ticket is for and hence the number of competitors on that route. Therefore, if the consolidators and the retail travel agents adopt the equilibrium behavior described here, then $M_1 = c_c + c_a$, which is independent of N. As result, $\frac{\partial M}{\partial N} = \frac{\partial M_2}{\partial N}$, which implies a pass-through rate of 1.

The result that the pass-through rate is 1 depends critically on the assumption that consolidators and retail travel agents compete in a homogenous product market. In reality, each consolidator and each retail travel agent may offer slightly different services and they may enjoy some ability to change their price-cost margins. In this case, the pass-through rate is strictly between 1 and 0 and we also have the result that $\mathrm{sgn}\left(\frac{\partial M}{\partial N}\right) = \mathrm{sgn}\left(\frac{\partial M_2}{\partial N}\right)$.

4.2 Empirical Models

The identification analysis above is for one ticket. Our data, however, is comprised of tickets sold by different airlines on different routes. Therefore, to estimate the marginal effect of market competition on price difference, which implies the relationship between price discrimination and market competition under our identification assumptions described above, we conduct regression analysis to control for other price influencing differences between different tickets. The empirical model is specified as follows:

(5). $$Price2_{ijkt} - Price1_{ijkt} = \beta_0 + \beta_1 N_{kt} + \beta_2 X_{ijkt} + \varepsilon_{ijkt},$$

Where $Price1_{ijkt}$ and $Price2_{ijkt}$ are the two prices from TravelSuperLink.com and FlyChina.com for the same ticket i by airline j on route k and dates t[68] and N_{kt} is the number of competitors, our measure for market competition.

X_{ijkt} is a set of variables for ticket characteristics and route- and airline-specific

---

[68] Each t corresponds to two dates, one for the US-China leg and the other for the China-US leg.

factors, which are likely to influence airlines' pricing decisions on one or both of the two prices we observe and hence the dependent variable in (5). It includes the following variables: Lag_i, $Advance_i$, $Duration_i$, $Weekend_i$, $Wait_i$, $Stops_i$, $Direct_i$, Hub_{jk}, $Distance_i$, $Load_{jkt}$, $Domo_{jkt}$, $Income_{kt}$ and $Chinese_{kt}$.[69] Some of these variables enter the regressions nonlinearly to capture the possible nonlinear effects of these variables on the dependent variable. $Advance$, $Duration$, $Weekend$, $Wait$, $Stops$ and $Direct$ are essentially ticket quality/restriction variables. In our data, we do not observe whether the ticket comes with an advance purchase requirement as in Stavins (2001) or when the cutoff date for the advance purchase requirement is. However, we observe the $Advance$ variable, the number of days in advance the ticket was purchased. As the shadow cost of capacity is different at different times, airlines charge different prices at different times. $Duration$ captures the stay restriction associated with the ticket. In the US domestic market, airlines offer discounts to tickets with a Saturday-night stayover requirement, as studied by both Stavins (2001) and Hernandez and Wiggins (2008). In the international market between the US and China, stay restriction is also an effective tool by the airlines to segment consumers as some consumers are willing and able to stay longer than others. $Weekend$ indicates travel restrictions associated with the ticket. Airlines often offer discounted tickets with the restriction that the tickets cannot be used for travel on certain days of a week or during certain time period. The travel restriction studied by Hernandez and Wiggins (2008) is similar to our $Weekend$ variable here. $Wait$, $Stops$ and $Direct$ are ticket quality variables. A trip with more stops

---

[69] The $Distance$ variable has subscript of i rather than k because two tickets on the same route can still have different travel distances due to different connecting airports.

and longer layover time is less convenient to the consumer and the consumer is compensated with a lower price. The rest of the variables are likely to influence the demand and/or the cost of the flight service and hence the ticket prices.

In addition, X_{ijkt} also includes airline fixed effects and two sets of monthly fixed effects for flight dates. Since all the tickets in our data are of round-trip, each ticket has both a US-China leg and a China-US leg. As seasonality and time trend are important determinants of ticket price, we use two sets of monthly dummies to control for them, one set for when the US-China leg took place and the other set for when the China-US leg took place.

4.3 Endogeneity

In estimating a pricing equation like (5), it is well known that the market competition variable, N_{kt}, might be endogenous, that is, N_{kt} might be correlated with the error term ε_{ijkt}. Any uncontrolled route or airline specific demand and/or supply factor that influence the ticket price and the number of airlines serving the route at the same time can cause such an endogeneity problem.

In the particular context of this study, however, the endogeneity problem is not likely to be severe, especially for observations from routes between a non-gateway airport in the US and China,[70] after controlling for many ticket price determinants including the full set of airline and monthly dummies. Suppose route k is a route between a non-gateway airport in the US and PEK in China. Whether airline j is able to serve this route depends on whether the airline offers flights between the non-gateway airport and one of its gateway airports, which are usually hubs of the airline, and whether the airline offers flights to China, which,

---

[70] Gateway airports are airports with international flights.

by definition, depart from one of its gateway airports. Both of these decisions are not likely to be determined by route k specific demand and supply conditions. First, whether the airline offers flights between the non-gateway airport and one of its gateway airports or hubs (and hence includes the non-gateway airport as part of its network) depends on the demand for flight services by passengers originating from (going to) this airport to (from) all airports on the airline's network, not just by passengers for route k. Similarly, whether the airline offers flights from gateway airports to China depends on the demand for flight services by passengers to (from) China from (to) all its airports in the US, not just passengers for route k. Therefore, the number of airlines that are able to serve route k is not likely to be correlated with time varying route k specific demand and/or supply conditions. However, many observations in our data come from routes between a gateway airport in the US and China. The argument here for the exogeneity of the market competition variable is weaker and hence the potential endogeneity problem is more severe for these observations.

We use the instrumental variable (IV) approach to address the potential endogeneity problem. In the context of this study, a valid IV needs to satisfy two conditions. First, it needs to be correlated with the potentially endogenous variable, that is, N_{kt}. Second, it needs to be uncorrelated with the uncontrolled route k specific demand and supply conditions. We use the population of the US end MSA of route k in the year when the ticket was purchased (POP_{kt}) as the IV variable for N_{kt}. As discussed above, whether an airline is able to serve route k depends on whether the airline offers flights between the local airport and one of its gateway (hub) airports. Airlines are more likely to offer flights from their hubs to airports with more population as more people demand more flight services. Therefore, POP_{kt} is likely to be correlated with N_{kt} for observations from routes between a non-gateway airport in the US and China. Airlines are also more likely to offer international flights to China from those gateway (hub) airports with larger population because again, more people demand more flight service. Therefore, POP_{kt} is also likely to be correlated with N_{kt} for observations from routes between a gateway airport in the US and China.

On the other hand, population is not likely to be correlated with the uncontrolled demand and/or supply factors for flight services to China, especially after we have already controlled for the number of Chinese-ethnic population in (5). Borenstein and Rose (1994), Hernandez and Wiggins (2008), Gerardi and Shapiro (2009) also use popultation as an IV for market competition variables in their studies.

5. Results

OLS and 2SLS are used to estimate (5). Since the relationships between ticket price difference and $Advance$ and $Duration$ are likely to be nonlinear and complex, a set of categorical dummy variables is used for these two variables in all specifications. This can be regarded as a non-parametric specification between the price difference and these two variables. The following categories are used for the $Advance$ variable: 0-7 days, 8-14 days, 15-21 days, 22-28 days, 29-35 days, 36-45 days, 46-60 days, 61-90 days, 91-120 days, 121-150 days, 151_180 days and over 181 days. For the $Duration$ variable, the following categories are used: 0-5 days, 6-13 days, 14-21 days, 22-30 days, 31-45 days, 46-60 days, 61-92 days, 93-120 days, 121-183 days and over 184 days. In addition, the relationship between the price difference and the *Duration* variable is specified to quadratic in all regressions. Finally, the standard errors used to compute the t-statistics are robust and clustered by route. This allows the error terms in (5) for observations from the same route to have arbitrary correlations.

Based upon the matching degree between the reporting itinerary from price data and the corresponding OAG itinerary, the 3,810 observations can be further classified into several groups[71]. There are 3,035 observations that are perfectly or partially but one-to-one[72] matched with OAG schedule data. Those observations are used as the main analysis. Another 775 observations can be partially, but one-to-many matched with schedule data. Robustness check is conducted by including those observations for regression.

5.1 Baseline Results

In the baseline regressions, we specify the relationship between price difference and the number of competitors to be linear. Estimation results for OLS and 2SLS are collected in Table 2, with selected results and test statistics from the first stage of the 2SLS reported in Table 3. Table 3 shows that the Angrist-Pischke χ^2 test rejects the hypothesis that the IV regression is under identified and the Angrist-Pischke F test rejects the hypothesis that the IV regression is weakly identified. As the OLS results may suffer from the endogeneity bias,

---

[71] Another 82 observations of the 3,892 ticket sample cannot be matched with the OAG schedule data at all, based upon the connection(s) information. Those observations are not included in any analysis due to the fact that the corresponding variables, such as, Wait and Stops cannot be created.

[72] One-to-one perfect match refers to the observations that one price observation can be perfectly matched and only matched with one schedule observation generated from OAG data in regards to itinerary; one-to-many partial match refers to the price observations that one price observation can be match with more than one schedule observations, and the connection information reported by price observations can be only partially matched with the schedule observations. The schedule among the matched itineraries with the shortest distance is kept for further variable creations.

we focus our discussion on the 2SLS results.

First of all, price difference increases in competition and the effect is statistically significant. Under the identification assumption laid out in section 4.1, this implies that the third-degree price discrimination in this market increases with competition. Stole (2007, pp. 2235-2236) compares the oligopoly third-degree price discrimination model by Holmes (1989) with the monopoly third-degree price discrimination model by Robinson (1933) and concludes that the effect of competition on price dispersion across markets (third-degree price discrimination) is ambiguous and depends on the cross-price elasticities. If the products are close substitutes or consumers are not brand loyal in both markets, competition will be fierce in each market. As a result, prices in both markets are lower and close to marginal cost and hence competition leads to less price discrimination. On the other hand, if consumers are brand loyal in one market but not in the other, then the price in the market where consumers are brand loyal will fall relatively less than that of the market where consumers are not brand loyal. As a result, competition increases price discrimination. Our result here is consistent with the latter case. When competition increases, airlines lower both the published fare and the agent fare because they enjoy lower market power, but they lower prices more for the group of travelers who purchase through the agents. Travelers who purchase at the published fare are travelers with high search costs. These travelers are likely to be business travelers or high-income leisure travelers, who are also more likely to be a member of a specific airline’s frequent flyer program. Therefore, these travelers are likely to be more brand loyal than travelers who purchase tickets through the agents. As a result, airlines have to compete more intensively for the group of travelers who purchase through agents by cutting price more and the price difference, or third-degree price discrimination, increases.

Second, all of the *Advance* categorical variables have a significant negative effect on price difference. This implies that the difference between the published fare and the agent fare is smaller when the ticket is purchased more than seven days in advance rather than less than seven days in advance. This is consistent with airlines increasing the published fares as the flight departure date is nearer. Third, the price difference tends to be larger when flights involved have a higher load factor. This is consistent with the airlines posting a higher published fare for routes that are more crowded. Fourth, the price difference is smaller for flights going through airports with slot controls. The costs of serving flights going through airports with slot controls may be particularly high and hence airlines are not willing to offer large discounts through agents for such tickets. Fifth, the price difference is larger if the layover time is longer. Longer layover time makes the travel experience less enjoyable. As a result, airlines may offer bigger discounts through agents to clear the inventory of such tickets. Sixth, the price difference is larger for travel during the peak season. The monthly dummy variables are included in all regressions. For brevity purpose, their coefficients are not reported in the table. Table 4 reports the average coefficient for the monthly dummies for the peak season and for the off-peak season separately. The Table 4 clearly shows that the price difference is larger for travel during peak season. This is consistent with the idea that airlines have more incentives to price discriminate during peak travel season, as we argued in section 4.1 above.

Some statistically insignificant results are also worth mentioning. First, the *Lag* dummy variables do not have a significant effect on price difference. This indicates that the small time difference between when the ticket was purchased and when we searched for the published fare is not likely to have a significant impact on our results. Also, all of the *Duration* categorical variables have an insignificant effect on price difference, indicating that airlines do not use stay restrictions to practice third-degree price discrimination in this market.

5.2 Main Results

In the baseline regressions, price discrimination is specified to be a parametric and linear function of the market competition variable. Recently, Dai, Liu and Serfes (2012) show that market competition has a non-monotone (non-linear) effect on price dispersion in the US airline market. They argue that this result could be driven by the fact that the effect of market competition on second-degree price discrimination is nonlinear. Hernandez and Wiggins (2008) also find evidence that the relationship between second-degree price discrimination and market competition is nonlinear in the US airline market. To consider this possibility for the third-degree price discrimination, instead of using the number of competitors variable as a continuous variable, we use a set of categorical dummy variables to measure market competition. We create 4 dummy variables, one for each quartile. For each dummy variable, the variable equals 1 if the number of competitors value falls into the corresponding quartile. The four quartiles are [0, 2]; [3, 5]; [6, 8] and [9, 11]. The dummy variable for the first quartile is omitted to avoid multicollinearity. Correspondingly, we also replace the population variable with its quartile dummy variables counterpart. Using quartile dummies can be thought as a less parametric approach to specify the relationship between price discrimination and market competition.

Regression results from this specification are reported in Tables 5—7. We find that price difference or third-degree price discrimination increases with market competition, but the effect is significant only when the competition increases from the 1st quartile to the 4th quartile. The changes are not statistically significant when market competition increases from the 1st quartile to the 2nd or the 3rd quartiles. This indicates that airlines change their price discrimination strategy only when there is a large change in the market competition condition. All other results essentially remain the same as those from the baseline regressions.

5.3 Robustness Check 1

Among the 3,035 observations used for main results analysis, there are 345 observations with a published fare lower than the agent fare. This might be due to the fact that, sometimes, especially during off-peak season when consolidators contract for a smaller number of tickets in bulk, travel agents run out of deeply discounted tickets. When that happens, agents can only sell consumers tickets at the published fare amount plus a service fee. Another reason could be consumers purchase the tickets from agents who do not work with consolidators. These imply that for these 345 tickets, airlines do not practice third-degree price discrimination. Including them in the sample may bias our estimates.

To check this, we analyzed repeat all of our regressions with a smaller sample where the 345 tickets are dropped. Results are reported in Tables 8—13. As we can see from the Tables, our results remain qualitatively the same and quantitatively similar.

5.4 Robustness Check 2

As mentioned above, there are 775 price observations can be partially and one-to-many matched with schedule observations in the sample. In order to check whether the results are consistent, we repeat all of the regressions with a bigger sample where the 775 price observations are included. The 775 price observations can be partially matched with the corresponding schedule observations, which means that the information reported by the consumers are mainly trustworthy, and the connection information cannot be perfectly matched with schedule information out of the possibility that the consumers skipped part of the connection information report for simplicity. Since the matching schedule is not unique, I picked the ones that match the itinerary the most and cover the shortest distance. First of all, the schedule information that matches the reported itinerary the most have the least number of connections, and therefore the distance will be the shortest among all of the matching observations[73]. Second of all, even if there are multiple matching itineraries that have the same number of stop(s) but different connection airport(s), the difference in distances across the itineraries is trivial[74]. Results are reported in Tables 14—19. It is shown that the third degree price discrimination increases as the market competition grows, and the results are statistically and economically significant. The other results remain qualitatively the same and quantitatively similar.

---

[73] For example, if a reporting ticket shows the itinerary as RDU-PEK, without connection information, the possible matching schedules, however, all have connection information, for example, RDU-DTW-PEK and RDU-DTW-NRT-PEK; the RDU-DTW-PEK is picked over the RDU-DTW-NRT-PEK.

[74] For example, if the reporting ticket shows the itinerary as PGV-CLT-PEK, but the possible matching schedules all have another connection in the US, for example: PGV-CLT-ORD-PEK (about 7,380 miles one way) and PGV-IAD-ORD-PEK (about 7,443 miles one way), the one with least distance (PGV-CLT-ORD-PEK) will be picked. The difference between the two itineraries is less than 1% of the average.

6. Conclusion

In this paper, we analyze how airlines exploit imperfect information and consumers' search costs as a fence to segment consumers and price discriminate in the US-China international flight market. Using a unique dataset collected from this market, we find large difference in prices for the same ticket from different ticket distribution channels and evidence that price discrimination increases with market competition. Thus, our paper contributes to the literature on the relationship between price discrimination and market competition in the airline industry in two ways. First, we offer new evidence of this relationship from an international flight market, while most of the previous studies focus on US domestic market. Second, we study price discrimination using search costs as a tool to segment consumers, while previous studies focus on price discrimination with respect to ticket qualities/restrictions.

Though our results are robust to changes in specifications and sample cleaning procedures, it is important to bear in mind the following caveats when interpreting our results. First, we do not observe the price airlines charge consolidators and hence we need to rely on an identification assumption to infer the relationship between price discrimination and competition from the relationship between price difference and competition. Though we believe our identification assumption is weak, it is still an assumption. Second, as our data spans only one year, we only have cross-sectional data on our instrumental variable, the population variable. As a result, we cannot control for route fixed effects in our regression analysis. Route fixed effects can control for many of the unobserved time-invariant route effects that are likely to influence airlines' pricing decisions. Controlling for these effects make the results less likely to suffer from the omitted variable bias problem. These are left for future research.

REFERENCES

Angrist, J. D. and J.-S. Pischke (2009): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Asplund, M., R. Eriksson and N. Strand (2008): “Price Discrimination in Oligopoly: Evidence from Swedish Newspapers,” *Journal of Industrial Economics*, 56, 2, 333-346.

Baye, M. and J. Morgan (2004): “Price Dispersion in the Lab and on the Internet: Theory and Evidence,” *RAND Journal of Economics*, 35, 3, 449-466.

Borenstein, S. (1985): “Price Discrimination in Free-Entry Markets,” *RAND Journal of Economics*, 16, 3, 380-397.

Borenstein, S. (1989): “Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry,” *RAND Journal of Economics*, 20, 3, 344-365.

Borenstein, S. and N. Rose (1994): “Competition and Price Dispersion in the U.S. Airline Industry,” *Journal of Political Economy*, 102, 4, 653-683.

Borenstein, S. (2011): “What Happened to Airline Market Power,” UC Berkeley Working Paper.

Borzekowski, R., R. Thomadsen and C. Taragin (2009): “Competition and Price Discrimination in the Market for Mailing Lists,” *Quantitative Marketing and Economics*, 7, 147-179.

Brueckner, J. K. (2001): “The Economics of Internaitonal Codesharing: An Anlysis of Airline Alliances,” *International Journal of Industrial Organization*, 19, 1475-1998.

Brueckner, J. K. (2003): “International Airfares in the Age of Alliances: The Effects of Codesharing and Antitrust Immunity,” *Review of Economics and Statistics*, 85, 1, 105-118.

Brueckner, J. K. and W. T. Whalen (2000): “The Price Effects of International Airline Alliances,” *Journal of Law and Economics*, 43, 503-545.

Busse, M. and M. Rysman (2005): “Competition and Price Discrimination in Yellow Pages Advertising,” *RAND Journal of Economics*, 36, 2, 378-390.

Clerides, S. and A. Michis (2006): “Market Concentration and Nonlinear Pricing: Evidence from Detergent Prices in Six Countries,” University of Cyprus Working Paper.

Dai, M., Q. Liu and K. Serfes (2012): “Is the Effect of Competition on Price Dispersion Non-Monotonic? Evidence from the U.S. Airline Industry,” *Review of Economics and Statistics*, forthcoming.

Dana, Jr., J. D. (1998): “Advance-Purchase Discounts and Price Discrimination in Competitive Markets,” *Journal of Political Economy*, 106, 2, 395-422.

Dana, Jr., J. D. (1999a): “Using Yield Management to Shift Demand When the Peak Time is Unknown,” *RAND Journal of Economics*, 30, 3, 456-474.

Dana, Jr., J. D. (1999b): “Equilibrium Price Dispersion under Demand Uncertainty: The Roles of Costly Capacity and Market Structure,” *RAND Journal of Economics*, 30, 4, 632-660.

Gerardi, K. and A. Shapiro (2009): “Does Competition Reduce Price Dispersion? New Evidence from the Airline Industry,” *Journal of Political Economy*, 117, 1, 1-37.

Giaume, S. and S. Guillou (2004): “Price Discrimination and Concentration in European Airline Markets,” *Journal of Air Transport Management*, 10, 5, 305-310.

Graddy, K. (1995): “Testing for Imperfect Competition at the Fulton Fish Market,” *RAND Journal of Economics*, 26, 75-92.

Hernandez, M. and S. Wiggins (2008): “Nonlinear Pricing and Market Concentration in the U.S. Airline Industry,” Texas A&M University Working Paper.

Holmes, T. (1989): “The Effects of Third-Degree Price Discrimination in Oligopoly,” *American Economic Review*, 79, 1, 244-250.

Hong, H. and M. Shum (2006): “Using Price Distributions to Estimate Search Costs,” *RAND Journal of Economics*, 37, 2, 257-275.

Hortacsu, A. and C. Syverson (2004): “Product Differentiation, Search Costs and Competition in the Mutual Fund Industry”, *Quarterly Journal of Economics*, 2004, 119, 403-456.

Katz, M. (1984): “Price Discrimination and Monopolistic Competition,” *Econometrica*, 52, 6, 1453-1471.

Orlov, E. (2011): “How Does the Internet Influence Price Dispersion? Evidence from the Airlines Industry,” *The Journal of Industrial Economics*, 59, 1, 21-37.

Robinson, J. (1933): *The Economics of Imperfect Competition.* Macmillan, London.

Rosenthal, R. (1980): “A Model in which an Increase in the Number of Sellers Leads to a Higher Price,” *Econometrica*, 48, 1575-1580.

Salop, S. (1977): “The Noisy Monopolist: Imperfect Information, Price Dispersion, and Price Discrimination,” *Review of Economic Studies*, 44, 393-406.

Shepard, A. (1991): “Price Discrimination and Retail Configuration,” *Journal of Political Economy*, 99, 30-53.

Stahl, D. (1989): “Oligopolistic Pricing with Sequential Consumer Search,” *American Economic Review*, 79, 700-712.

Stavins, J. (2001): “Price Discrimination in the Airline Market: The Effect of Market Concentration,” *The Review of Economics and Statistics*, 83, 200-202.

Stole, L. (2007): “Price Discrimination and Competition,” In Mark Armstrong and Robert Porter, eds. *Handbook of Industrial Organization*, Vol. 3, pp. 2221-99. San Diego CA: Elsevier Science Publishers.

Tirole, J. (1988): “Price Discrimination,” In J. Tirole, *The Theory of Industrial Organization*, pp. 133-163. Cambridge, Massachusetts and London, England: The MIT Press.

Transportation Group Internatinoal, LC (2002): “Travel Agents Access to Airline Fares,” *A Research Report Prepared for the National Commission to Ensure Consumer Information and Choice in the Airline Industry.*

Varian, H. (1980): “A Model of Sales,” *American Economic Review*, 70, 651-659.

Yang, H. and L. Ye (2008): “Nonlinear Pricing, Contract Variety, and Competition,” *Theoretical Economics*, 31, 1, 123-153.

Tabel II-1: Summary Statistics[a]

Variable	Mean	Std Dev	Minimum	Maximum
Price1	1212.13	255.867885	628	2966
Price2	1779.64	1020.75	598	12520
Price2-Price1	567.513385	917.966434	-450	10945
Lag	0.8651079	0.6929134	0	2
USorigin	0.9326824	0.2506034	0	1
Hub	0.3137205	0.4640639	0	1
AA	0.3437821	0.4750305	0	1
DL	0.3024152	0.4593631	0	1
UA	0.3093525	0.4622861	0	1
Chinese_Airlines	0.0305755	0.1721868	0	1
Other_Airlines	0.0138746	0.1169856	0	1
Advance	44.8257965	29.8966002	2	184
Duration	51.3496917	39.9988405	2	323
PEK	0.4807297	0.4996927	0	1
Weekend	0.9231757	0.7590844	0	2
Distance	7133.52	629.311756	4768	11613.5
Wait	229.16277	207.403318	0	870
Stops	0.8238695	0.5770191	0	3
Slot	0.0187564	0.1356811	0	1
Direct	0.2667009	0.4422915	0	1
Comp	6.2792909	2.4352531	0	11
Load	0.8595807	0.0727879	0.4857218	0.9726448
Domo	0.2487113	0.1251056	0	0.5868634
Population	6082.2	5794.32	7.71	19015.9
Chinese	155.4405	225.827902	0	657.203
Income	3.0960136	0.6325249	1.5254	7.2415

[a] Notes: See variable definitions in Section 3. Income is in $10,000s. Population and Chinese are in 1,000s. Number of Observations is 3,892.

Table II-2: Baseline Regression Results[a]

	OLS		2SLS	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**
N	12.1571	0.93	79.9146**	2.25
=1 if Weekend=0	-86.9017*	-1.88	-83.9042*	-1.79
=1 if Weekend=1	-77.9342*	-1.75	-76.2544*	-1.71
Wait	0.3847***	3.17	0.5434***	3.81
=1 if Stops=0	142.1738**	2.18	146.4779**	2.08
=1 if Stops=1	206.3870**	1.98	206.7512**	2.00
=1 if Lag=0	-43.0701	-0.78	-45.2281	-0.83
=1 if Lag=1	-46.8926	-1.02	-50.4554	-1.12
Distance	0.6506	1.51	0.8031*	1.84
Distance^2	-0.0000*	-1.73	-0.0001**	-2.08
Slot	-191.5630**	-2.38	-374.1853***	-2.87
Load	1,312.3871***	3.53	1,402.5486***	3.52
Domo	398.7054	1.29	786.0831**	2.27
Income	-8.8468	-0.37	-38.7514	-1.02
Chinese	0.1403	0.84	-0.1973	-0.81
Hub	37.3340	0.38	-80.3071	-0.67
USorigin	166.9878**	2.31	153.2983**	1.98
PEK	177.5481***	3.45	159.0203***	3.07
=1 if 8<=Advance<=14	-393.4750***	-3.14	-405.0233***	-3.20
=1 if 15<=Advance<=21	-531.6675***	-4.34	-543.1690***	-4.46
=1 if 22<=Advance<=28	-456.9218***	-3.48	-473.6392***	-3.58
=1 if 29<=Advance<=35	-476.8013***	-3.71	-487.0731***	-3.72
=1 if 36<=Advance<=45	-483.1238***	-4.09	-492.4270***	-4.14
=1 if 46<=Advance<=60	-454.0116***	-3.79	-466.5301***	-3.84
=1 if 61<=Advance<=90	-457.4580***	-3.73	-467.1594***	-3.78
=1 if 91<=Advance<=120	-774.5218***	-5.88	-790.0073***	-5.92
=1 if 121<=Advance<=150	-902.8551***	-5.65	-921.4361***	-5.92
=1 if 151<=Advance<=180	-885.7822***	-3.56	-944.3471***	-3.70
=1 if 181<=Advance	-624.3379***	-3.94	-714.3322***	-4.82
=1 if 6<=Duration<=13	-198.1285	-0.78	-197.1092	-0.78
=1 if 14<=Duration<=21	-161.9490	-0.66	-160.4572	-0.66
=1 if 22<=Duration<=30	-155.1459	-0.64	-149.5815	-0.62
=1 if 31<=Duration<=45	-177.1687	-0.73	-182.1128	-0.75
=1 if 46<=Duration<=60	-196.8009	-0.80	-199.8780	-0.81
=1 if 61<=Duration<=92	-60.2484	-0.25	-65.3786	-0.27
=1 if 93<=Duration<=120	217.0371	0.81	211.6524	0.79
=1 if 121<=Duration<=183	-7.1347	-0.03	-3.2650	-0.01
=1 if 184<=Duration	-550.5929	-1.47	-466.9006	-1.28
Constant	-2,681.4544	-1.56	-3,375.2103*	-1.94
Adjusted R^2	0.355		0.346	

[a] Number of observations is 3,035. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-3: Selected Results from First-stage Regressions[a]

	N **Estimates**	**t-stat**
Population	0.0003***	5.42
AP χ^2 statistic	30.08	
p-value	0.0000	
AP F test statistic	29.33	
Adjusted R^2	0.752	

[a] Number of observations is 3,035. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-4: Average of Estimated Coefficients for Monthly Dummy Variables[a]

OLS		**2SLS**	
peak	**off-peak**	**peak**	**off-peak**
-24.76	-158.67	-56.67	-205.42

[a] This is a summary of the parameter estimates on monthly dummies. We group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Table II-5: Main Regression Results[a]

	OLS		2SLS	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**
=1 if N=3, 4, 5	59.7104	0.81	110.6943	0.67
=1 if N=6, 7, 8	97.3047	1.11	6.0072	0.03
=1 if N=9, 10, 11	81.3425	0.70	458.4874*	1.88
=1 if Weekend=0	-88.4283*	-1.95	-69.1832	-1.42
=1 if Weekend=1	-79.3455*	-1.81	-68.5776	-1.54
Wait	0.3697***	3.04	0.5336***	3.66
=1 if Stops=0	136.0581**	2.00	194.9323**	2.39
=1 if Stops=1	204.9363*	1.91	181.9573	1.58
=1 if Lag=0	-43.4219	-0.79	-36.0257	-0.67
=1 if Lag=1	-47.9904	-1.05	-35.1841	-0.78
Distance	0.6870	1.59	0.1467	0.28
Distance^2	-0.0001*	-1.79	-0.0000	-0.40
Slot	-174.4207**	-2.19	-364.3033***	-2.78
Load	1,319.2727***	3.62	1,282.9114***	3.05
Domo	376.4193	1.25	752.6815**	1.99
Income	-7.9184	-0.33	-8.5396	-0.33
Chinese	0.1936	1.12	-0.1238	-0.59
Hub	37.7310	0.38	32.4528	0.27
USorigin	167.4320**	2.31	153.0400**	1.99
PEK	178.9163***	3.47	159.0765***	2.85
=1 if 8<=Advance<=14	-389.6646***	-3.06	-434.4813***	-3.40
=1 if 15<=Advance<=21	-529.1633***	-4.28	-566.3847***	-4.62
=1 if 22<=Advance<=28	-455.1388***	-3.44	-479.5524***	-3.65
=1 if 29<=Advance<=35	-473.0023***	-3.62	-516.0797***	-3.90
=1 if 36<=Advance<=45	-479.7075***	-4.01	-521.4940***	-4.33
=1 if 46<=Advance<=60	-450.5401***	-3.73	-495.5634***	-4.01
=1 if 61<=Advance<=90	-454.0686***	-3.67	-498.1945***	-3.93
=1 if 91<=Advance<=120	-768.8344***	-5.76	-832.1983***	-6.06
=1 if 121<=Advance<=150	-899.5658***	-5.62	-920.4098***	-5.58
=1 if 151<=Advance<=180	-876.7592***	-3.48	-960.2044***	-3.36
=1 if 181<=Advance	-598.8914***	-3.36	-885.7175***	-6.33
=1 if 6<=Duration<=13	-204.2988	-0.79	-154.0867	-0.62
=1 if 14<=Duration<=21	-168.8994	-0.68	-109.4275	-0.45
=1 if 22<=Duration<=30	-163.1148	-0.67	-92.5599	-0.39
=1 if 31<=Duration<=45	-182.9237	-0.75	-132.2677	-0.56
=1 if 46<=Duration<=60	-202.1799	-0.81	-158.8616	-0.65
=1 if 61<=Duration<=92	-65.6843	-0.27	-24.0485	-0.10
=1 if 93<=Duration<=120	211.8067	0.78	251.7107	0.95
=1 if 121<=Duration<=183	-13.9635	-0.06	42.1736	0.17
=1 if 184<=Duration	-564.2634	-1.51	-466.1143	-1.21
Constant	-2,818.1585	-1.65	-1,032.5222	-0.50
Adjusted $\bar{R}^2$	0.355		0.334	

[a] Number of observations is 3,035. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-6: Selected Results from First-stage Regressions with Categorical N[a]

	=1 if N=5, 6		=1 if N=7, 8		=1 if N=9, 10, 11	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**	**Estimates**	**t-stat**
Population 2	0.5530***	4.38	0.0802	1.23	-0.0266	-1.04
Population 3	0.2707*	1.97	0.3538***	4.13	0.0377	1.16
Population 4	0.0715	0.49	0.0518	0.58	0.5171***	7.87
AP χ^2 statistic	20.92		18.13		68.77	
p-value	0.0000		0.0000		0.0000	
AP F test statistic	20.38		17.66		66.99	
Adjusted R^2	0.408		0.275		0.665	

[a] Number of observations is 3,035. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions. Population 2--4 are 2nd, 3rd and fourth quartile dummies for the Population variable.

Table II-7: Average of Estimated Coefficients for Monthly Dummy Variables[a]

OLS		2SLS	
peak	**off-peak**	**peak**	**off-peak**
-24.47	-158.43	-7.21	-140.40

[a] This is a summary of the parameter estimates on monthly dummies. We group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Table II-8: Robustness Check 1: Baseline Regression Results[a]

	OLS		2SLS	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**
N	10.1628	0.73	79.9292**	1.97
=1 if Weekend=0	-69.7996	-1.46	-69.8943	-1.45
=1 if Weekend=1	-71.1427	-1.56	-69.6188	-1.52
Wait	0.4388***	3.46	0.6011***	3.97
=1 if Stops=0	167.7213**	2.57	164.7085**	2.55
=1 if Stops=1	233.9607**	2.04	227.4050**	2.08
=1 if Lag=0	-51.4134	-0.76	-54.4166	-0.81
=1 if Lag=1	-61.8253	-1.18	-66.5393	-1.28
Distance	0.6589	1.33	0.8072	1.63
Distance^2	-0.0001	-1.56	-0.0001*	-1.88
Slot	-102.5431	-1.16	-290.6923**	-1.99
Load	1,423.5657***	3.42	1,540.6084***	3.40
Domo	309.7596	0.94	707.1148*	1.91
Income	-4.8151	-0.18	-33.6460	-0.84
Chinese	0.1912	1.04	-0.1629	-0.58
Hub	51.8915	0.50	-70.9190	-0.53
USorigin	207.9869**	2.39	192.5614**	2.10
PEK	168.1276***	2.99	149.1284**	2.51
=1 if 8<=Advance<=14	-428.8953***	-3.17	-443.2101***	-3.24
=1 if 15<=Advance<=21	-565.3340***	-4.29	-577.6346***	-4.41
=1 if 22<=Advance<=28	-491.0523***	-3.49	-509.9630***	-3.61
=1 if 29<=Advance<=35	-514.8867***	-3.70	-526.3725***	-3.71
=1 if 36<=Advance<=45	-516.9673***	-4.13	-524.4018***	-4.15
=1 if 46<=Advance<=60	-500.4163***	-3.95	-513.1599***	-3.99
=1 if 61<=Advance<=90	-476.7558***	-3.70	-487.5212***	-3.75
=1 if 91<=Advance<=120	-687.2472***	-4.69	-700.4587***	-4.69
=1 if 121<=Advance<=150	-917.8696***	-5.65	-941.6885***	-5.91
=1 if 151<=Advance<=180	-973.9025***	-3.84	-1,026.6860***	-3.93
=1 if 181<=Advance	-706.6805***	-4.31	-801.6846***	-5.16
=1 if 6<=Duration<=13	-335.2379	-1.14	-320.0858	-1.08
=1 if 14<=Duration<=21	-306.1243	-1.08	-292.2749	-1.02
=1 if 22<=Duration<=30	-296.8361	-1.06	-278.1045	-0.98
=1 if 31<=Duration<=45	-314.5669	-1.13	-303.9461	-1.08
=1 if 46<=Duration<=60	-314.7001	-1.10	-304.6050	-1.06
=1 if 61<=Duration<=92	-156.7268	-0.55	-146.8705	-0.51
=1 if 93<=Duration<=120	175.2588	0.57	184.7222	0.60
=1 if 121<=Duration<=183	-169.1694	-0.58	-145.1545	-0.50
=1 if 184<=Duration	-1,025.0530**	-2.38	-899.2096**	-2.17
Constant	-2,232.5246	-1.15	-2,968.3197	-1.50
Adjusted R^2	0.374		0.365	

[a] Number of observations is 2,690. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-9: Selected Results from First-stage Regressions[a]

	N Estimates	t-stat
Population	0.0003***	5.47
AP χ^2 statistic	30.75	
p-value	0.0000	
AP F test statistic	29.89	
Adjusted R^2	0.754	

[a] Number of observations is 2,690. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-10: Average of Estimated Coefficients for Monthly Dummy Variables[a]

OLS		**2SLS**	
peak	**off-peak**	**peak**	**off-peak**
-226.38	-367.45	-248.33	-403.63

[a] This is a summary of the parameter estimates on monthly dummies. We group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Table II-11: Robustness Check 1: Main Regression Results[a]

	OLS		2SLS	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**
=1 if N=3, 4, 5	51.2888	0.65	113.4687	0.67
=1 if N=6, 7, 8	89.9728	0.94	-42.8752	-0.19
=1 if N=9, 10, 11	52.9869	0.42	439.5135*	1.67
=1 if Weekend=0	-71.9072	-1.54	-54.2676	-1.09
=1 if Weekend=1	-73.0104	-1.62	-60.5370	-1.33
Wait	0.4170***	3.32	0.5968***	3.81
=1 if Stops=0	159.2968**	2.30	237.6389***	2.58
=1 if Stops=1	233.1978*	1.96	223.3028*	1.71
=1 if Lag=0	-52.1073	-0.77	-45.9741	-0.69
=1 if Lag=1	-63.2136	-1.22	-50.0911	-0.98
Distance	0.7227	1.44	0.0311	0.05
Distance^2	-0.0001*	-1.65	-0.0000	-0.16
Slot	-78.8044	-0.91	-265.7053*	-1.73
Load	1,430.5282***	3.50	1,407.7615***	2.92
Domo	272.8343	0.85	689.0452*	1.71
Income	-4.2020	-0.16	-4.4094	-0.15
Chinese	0.2532	1.33	-0.0831	-0.34
Hub	52.0583	0.49	59.1885	0.44
USorigin	208.9304**	2.40	188.6475**	2.14
PEK	170.4924***	3.03	145.8974**	2.21
=1 if 8<=Advance<=14	-423.3614***	-3.09	-470.1868***	-3.44
=1 if 15<=Advance<=21	-561.7299***	-4.23	-597.2907***	-4.54
=1 if 22<=Advance<=28	-488.0938***	-3.44	-512.3614***	-3.65
=1 if 29<=Advance<=35	-508.8984***	-3.59	-563.3938***	-3.98
=1 if 36<=Advance<=45	-511.8233***	-4.03	-558.3354***	-4.40
=1 if 46<=Advance<=60	-495.4181***	-3.88	-542.0645***	-4.17
=1 if 61<=Advance<=90	-472.1495***	-3.65	-515.6037***	-3.91
=1 if 91<=Advance<=120	-679.2772***	-4.60	-748.8799***	-4.89
=1 if 121<=Advance<=150	-912.5804***	-5.62	-939.8694***	-5.46
=1 if 151<=Advance<=180	-963.1247***	-3.77	-1,037.5301***	-3.54
=1 if 181<=Advance	-671.8915***	-3.63	-979.0462***	-6.63
=1 if 6<=Duration<=13	-346.2082	-1.16	-265.9152	-0.90
=1 if 14<=Duration<=21	-317.7965	-1.11	-233.0870	-0.82
=1 if 22<=Duration<=30	-310.0195	-1.09	-210.1972	-0.75
=1 if 31<=Duration<=45	-324.8821	-1.16	-247.0021	-0.89
=1 if 46<=Duration<=60	-323.5032	-1.12	-259.4675	-0.91
=1 if 61<=Duration<=92	-166.8889	-0.58	-95.4778	-0.33
=1 if 93<=Duration<=120	165.8044	0.54	230.5953	0.76
=1 if 121<=Duration<=183	-181.0556	-0.62	-100.2403	-0.35
=1 if 184<=Duration	-1,038.5506**	-2.39	-990.9087**	-2.32
Constant	-2,459.9762	-1.27	-166.4522	-0.07
Adjusted R^2	0.374		0.348	

[a] Number of observations is 2,690. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-12: Selected Results from First-stage Regressions with Categorical N[a]

	=1 if N=5, 6		=1 if N=7,8		=1 if N=9,10,11	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**	**Estimates**	**t-stat**
Population 2	0.5289***	3.95	0.0818	1.19	-0.0264	-1.00
Population 3	0.2501*	1.73	0.3447***	3.90	0.0462	1.34
Population 4	0.0402	0.26	0.0506	0.56	0.5269***	8.03
AP χ^2 statistic	17.31		17.30		70.53	
p-value	0.0000		0.0000		0.0000	
AP F test statistic	16.81		16.80		68.50	
Adjusted R^2	0.407		0.281		0.665	

[a] Number of observations is 2,690. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions. Population 2--4 are 2nd, 3rd and fourth quartile dummies for the Population variable.

Table II-13: Average of Estimated Coefficients for Monthly Dummy Variables[a]

OLS		**2SLS**	
peak	**off-peak**	**peak**	**off-peak**
-226.25	-367.33	-214.98	-355.24

[a] This is a summary of the parameter estimates on monthly dummies. We group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Table II-14: Robustness Check 2: Baseline Regression Results[a]

	OLS		2SLS	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**
N	16.8154	1.44	83.8510***	2.63
=1 if Weekend=0	-82.4108**	-2.18	-81.2343**	-2.13
=1 if Weekend=1	-70.5226**	-2.03	-68.6933*	-1.96
Wait	0.3962***	2.96	0.5617***	3.61
=1 if Stops=0	-112.9147	-1.16	-111.5929	-1.20
=1 if Stops=1	-67.4293	-0.66	-75.8854	-0.74
=1 if Lag=0	-30.2061	-0.64	-36.3572	-0.78
=1 if Lag=1	-24.7373	-0.68	-28.5759	-0.78
Distance	0.3643	1.13	0.4615	1.36
Distance^2	-0.0000	-1.36	-0.0000	-1.61
Slot	-145.5666	-1.51	-316.0605**	-2.33
Load	1,038.4389***	3.26	1,085.9074***	3.24
Domo	434.1275	1.63	774.6067***	2.63
Income	-5.7807	-0.24	-37.2302	-0.97
Chinese	0.1878	1.09	-0.1481	-0.68
Hub	27.9946	0.33	-72.6152	-0.75
USorigin	60.3793	0.51	55.8790	0.46
PEK	161.7332***	3.49	152.1106***	3.46
=1 if 8<=Advance<=14	-407.3119***	-4.09	-415.4004***	-4.17
=1 if 15<=Advance<=21	-536.9450***	-5.11	-546.5253***	-5.29
=1 if 22<=Advance<=28	-521.5777***	-4.48	-536.6143***	-4.57
=1 if 29<=Advance<=35	-508.5030***	-4.55	-513.9078***	-4.52
=1 if 36<=Advance<=45	-524.9557***	-5.10	-527.2128***	-5.15
=1 if 46<=Advance<=60	-475.4953***	-4.47	-483.1982***	-4.53
=1 if 61<=Advance<=90	-495.5534***	-4.72	-499.7332***	-4.79
=1 if 91<=Advance<=120	-722.8090***	-5.98	-732.7281***	-6.03
=1 if 121<=Advance<=150	-823.8383***	-5.62	-831.4219***	-5.80
=1 if 151<=Advance<=180	-859.8539***	-3.47	-911.3975***	-3.60
=1 if 181<=Advance	-604.9794***	-4.21	-696.9330***	-5.31
=1 if 6<=Duration<=13	-131.3892	-0.72	-140.5364	-0.76
=1 if 14<=Duration<=21	-110.2125	-0.63	-119.0409	-0.67
=1 if 22<=Duration<=30	-118.4296	-0.69	-123.9623	-0.71
=1 if 31<=Duration<=45	-86.3092	-0.50	-100.8037	-0.58
=1 if 46<=Duration<=60	-132.7730	-0.75	-146.4269	-0.82
=1 if 61<=Duration<=92	-9.1741	-0.05	-24.2763	-0.14
=1 if 93<=Duration<=120	219.8979	1.11	212.4087	1.07
=1 if 121<=Duration<=183	1.6816	0.01	-2.8657	-0.02
=1 if 184<=Duration	-36.6197	-0.10	-26.0433	-0.07
Constant	-1,442.5293	-1.11	-1,945.0122	-1.44
Adjusted R^2	0.310		0.300	

[a] Number of observations is 3,810. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-15: Selected Results from First-stage Regressions[a]

	N	
	Estimates	**t-stat**
Population	0.0003***	5.04
AP χ^2 statistic	25.92	
p-value	0.0000	
AP F test statistic	25.39	
Adjusted R^2	0.743	

[a] Number of observations is 3,810. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-16: Average of Estimated Coefficients for Monthly Dummy Variables[a]

OLS		2SLS	
peak	**off-peak**	**peak**	**off-peak**
105.32	-45.83	78.03	-87.21

[a] This is a summary of the parameter estimates on monthly dummies. We group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Table II-17: Robustness Check 2: Main Regression Results[a]

	OLS		2SLS	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**
=1 if N=3, 4, 5	107.6077*	1.72	141.6898	1.10
=1 if N=6, 7, 8	148.0981**	2.00	-52.9196	-0.33
=1 if N=9, 10, 11	143.7145	1.37	485.8502**	2.55
=1 if Weekend=0	-83.4902**	-2.27	-62.3870	-1.59
=1 if Weekend=1	-72.8003**	-2.14	-64.1968*	-1.85
Wait	0.3838***	2.90	0.4958***	3.05
=1 if Stops=0	-117.7068	-1.21	-61.2302	-0.66
=1 if Stops=1	-68.4866	-0.67	-98.8188	-0.92
=1 if Lag=0	-30.8458	-0.66	-20.7409	-0.45
=1 if Lag=1	-26.3882	-0.73	-11.4426	-0.31
Distance	0.3826	1.20	-0.0773	-0.23
Distance^2	-0.0000	-1.43	0.0000	0.09
Slot	-129.4136	-1.36	-296.9587**	-2.48
Load	1,044.1535***	3.31	1,036.8451***	2.91
Domo	437.7152*	1.68	646.2344**	1.98
Income	-5.1091	-0.21	-0.5892	-0.03
Chinese	0.2558	1.37	-0.1185	-0.65
Hub	27.0859	0.32	49.9899	0.52
USorigin	59.7729	0.50	42.4173	0.35
PEK	161.6084***	3.45	144.4785***	3.03
=1 if 8<=Advance<=14	-406.3468***	-4.03	-437.4472***	-4.29
=1 if 15<=Advance<=21	-536.8958***	-5.07	-565.9560***	-5.28
=1 if 22<=Advance<=28	-521.7193***	-4.44	-541.9714***	-4.62
=1 if 29<=Advance<=35	-505.2578***	-4.44	-555.2468***	-4.69
=1 if 36<=Advance<=45	-523.3582***	-5.01	-561.6409***	-5.28
=1 if 46<=Advance<=60	-474.1342***	-4.42	-508.1803***	-4.62
=1 if 61<=Advance<=90	-494.2418***	-4.66	-531.2758***	-4.87
=1 if 91<=Advance<=120	-719.6465***	-5.89	-776.7711***	-6.05
=1 if 121<=Advance<=150	-821.3439***	-5.58	-835.5273***	-5.37
=1 if 151<=Advance<=180	-850.3599***	-3.34	-964.3304***	-3.18
=1 if 181<=Advance	-580.1200***	-3.42	-897.0283***	-7.46
=1 if 6<=Duration<=13	-136.0608	-0.73	-99.3511	-0.54
=1 if 14<=Duration<=21	-116.0806	-0.66	-78.0319	-0.45
=1 if 22<=Duration<=30	-124.9742	-0.72	-81.2387	-0.47
=1 if 31<=Duration<=45	-90.4607	-0.51	-53.4772	-0.31
=1 if 46<=Duration<=60	-136.2525	-0.76	-120.8291	-0.69
=1 if 61<=Duration<=92	-13.7850	-0.08	5.7583	0.03
=1 if 93<=Duration<=120	216.4580	1.08	235.9830	1.20
=1 if 121<=Duration<=183	-5.3024	-0.03	22.7257	0.12
=1 if 184<=Duration	-52.7666	-0.14	32.2897	0.09
Constant	-1,536.6582	-1.21	-11.8970	-0.01
Adjusted R^2	0.310		0.279	

[a] Number of observations is 3,810. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions.

Table II-18: Selected Results from First-stage Regressions with Categorical N[a]

	=1 if N=5, 6		=1 if N=7,8		=1 if N=9, 10, 11	
VARIABLES	**Estimates**	**t-stat**	**Estimates**	**t-stat**	**Estimates**	**t-stat**
Population 2	0.6317***	5.55	0.0265	0.57	-0.0371*	-1.71
Population 3	0.2621**	2.07	0.3854***	5.17	0.0297	1.05
Population 4	0.0621	0.45	0.0900	1.13	0.5135***	7.34
AP χ^2 statistic	30.09		30.39		59.85	
p-value	0.0000		0.0000		0.0000	
AP F test statistic	29.46		29.76		58.60	
Adjusted R^2	0.418		0.269		0.670	

[a] Number of observations is 3,810. Standard errors used to compute the t-statistics are robust and clustered by route. Statistical significance at the 1%, 5%, and 10% level are denoted by ***, **, and *, respectively. Airline and monthly dummies are included in both regressions. Population 2--4 are 2nd, 3rd and fourth quartile dummies for the Population variable.

Table II-19: Average of Estimated Coefficients for Monthly Dummy Variables[a]

OLS		2SLS	
peak	**off-peak**	**peak**	**off-peak**
105.95	-45.16	146.08	-5.37

[a] This is a summary of the parameter estimates on monthly dummies. We group the monthly dummies into two groups, one is peak and the other is non-peak. This table reports the average parameter estimate for the non-peak dummies, and the average parameter estimate for the peak dummies. Peak travel period is defined as travel (either China-US direction or US-China direction) in January, May, June, July, August and December, which correspond to winter and summer vacation time.

Chapter 3

When a Negative Externality Turns Positive: An Analysis of Flue-gas Desulfurization By-products

1. Introduction

Sulfur dioxide (SO_2) is a gaseous pollutant emitted during the electricity generating process at coal fired electric plants. It is a precursor to acid rain, and it is associated with increased morbidity and mortality risk in humans (Ozkaynak & Spengler, 1985). Methods for removing, or scrubbing, SO_2 from electricity plants' flue gas emissions involve using an alkaline sorbent material to neutralize the SO_2. In 1981, approximately 63% of the total SO_2 scrubbers installed in the U.S. used limestone as the primary sorbent material (Henzel et al., 1981). Overtime, the consistently low price of limestone in comparison to other potential sorbent materials has insured limestone's position as the predominant SO_2 scrubbing material used in the U.S. As of 2010, 61% of the SO_2 scrubbers in operation relied on limestone as the primary sorbent material, and nearly all of the scrubbers using limestone produce a synthetic gypsum by-product.

As early as 1984, the U.S. Environmental Protection Agency (EPA) documented the potential marketability of flue-gas desulfurization (FGD) synthetic gypsum by-product. In a marketing study prepared for the EPA, O'brien et al. (1984) estimated that electricity generating plants have the ability to market their gypsum by-product up to 500 miles away when selling to cement plants, and up to 250 miles away when selling to wallboard manufacturers. The difference in marketability between the cement and wallboard markets is due to the fact that wallboard manufacturers typically vertically integrate the gypsum mining process and the wallboard production process. Wallboard manufacturers therefore control the source of their gypsum and in turn have a lower cost for the gypsum input.

The O'brien et al. (1984) marketing study was a case study based on 14 individual power plants that were operating limestone fed SO_2 scrubbers during 1984. The study found that all 14 power plants could reduce cost by marketing their by-product gypsum, and between 6 and 12 of the 14 plants could successfully market all of their synthetic gypsum. From 1985 to 2005, however, the realized growth in the synthetic gypsum market was much more subdued in comparison to the growth predicted by the EPA's case study. As of 1990, there were only three power plants successfully marketing their synthetic gypsum by-product. Post 1990 there has been an upward trend in the number of power plants selling gypsum, but the growth was generally slow through 2005. From 2005 until 2010 the FGD gypsum market size doubled from 30 electricity plants selling gypsum in 2005 to 63 gypsum sellers in 2010.

Overall, growth in the FGD by-product gypsum market has been considerably slower than the growth foreshadowed in the O'brien et al. (1984) case study. This is likely due to the fact that the EPA study assumed no competition and did not account for additional capital costs required to employ by-product gypsum in the production of wallboard. The more rapid growth in the FGD by-product market post 2005 is due to power plants increased retrofitting of SO_2 scrubbers in order to comply with the increased SO_2 emission reduction mandates slated to begin in 2010 (see, for example, USGS, 2012b, CAIR, 2005, and CAAA, 1990).

Our analysis finds that the development of a market for FGD gypsum has provided incentives for plants to increase the operating efficiency of their SO_2 scrubbers thereby providing significant reductions in total SO_2 emissions. Results indicate that the marketing of synthetic gypsum is associated with a 3,494 ton reduction in annual boiler SO_2 emissions. The aggregate external benefits associated with synthetic gypsum production from 1990 to 2010 are estimated between \$147.1 billion and \$350.6 billion (2011 \$). These aggregate benefits account for between 0.9% and 2.1% of the aggregate benefits of the U.S. Clean Air Act (CAA) calculated over the same time period (U.S. EPA 2011).

The remainder of this paper is organized as follows. Section 2 provides a theoretical model explaining electricity plants' ability to market their FGD gypsum by-product. The profit maximization decision of an electricity plant with a contract to sell by-product gypsum is compared to the profit maximization decision in the absence of a gypsum sales contract. The theoretical model from section 2 suggests that electricity generating plants with gypsum contracts will operate their scrubbers more efficiently and reduce their SO_2 emissions. Section 3 presents an overview of the EPA and EIA datasets that are used for analysis, and section 4 provides empirical evidence that power plants with gypsum contracts do indeed reduce their SO_2 emissions. Finally, section 5 concludes.

2. Theory

Synthetic gypsum suitable for wallboard manufacturing is produced as a by-product during the SO_2 scrubbing process for scrubbers using lime and limestone as a sorbent material. SO_2 scrubbers using limestone or lime as the primary sorbent materials account for roughly 79% of the total SO_2 scrubbers operating in the U.S. as of year 2010. Limestone scrubbers generally have a higher operating cost, because limestone is chemically less reactive and must first be ground to a smaller size in order to allow for decomposition in water (Srivastava, 2000). Despite having a higher operating cost, limestone is the most widely used sorbent material (61%) because the price of limestone is considerably lower than the price of lime. Figure 1 depicts the price of lime and limestone for the time period 1985 to 2010. As Figure 1 illustrates, the price of limestone is roughly 9% of the price of lime, and is consistently between 8% and 10% of the price of lime for all years during the 1985 to 2010 timeframe. Due to the consistent extreme differences in material costs, lime scrubbing procedures must achieve considerable savings in operating costs through technological improvements or improvements in reactivity such as magnesium enhancing in order to compete with limestone as a cost-effective sorbent alternative (Srivastava, 2000). The theoretical model and empirical analysis that follow focus solely on limestone scrubbers, because these scrubbers account for over 91% of the observations in our dataset that have contracts to sell FGD gypsum. Because lime is a more reactive sorbent material, including lime scrubbers in the estimation of the SO_2 emission function may bias the estimated impact of sorbent use on SO_2 emissions.

According to Flagan and Seinfeld (1988), the overall chemical reaction produced during the limestone ($CaCO_3$) wet scrubbing process is given by the following:

$$(1).\quad CaCO_3 + SO_2 + 2H_2O \rightarrow CaSO_3 \cdot 2H_2O + CO_2 .$$

Gypsum ($CaSO_4 \cdot 2H_2O$) is then produced from the calcium sulfite ($CaSO_3 \cdot 2\ H_2O$) by-product of limestone scrubbing using the method of forced oxidation. The chemical reaction of forced oxidation is given by the following:

(2). $CaSO_3 \cdot 2H_2O + \frac{1}{2}O_2 \rightarrow CaSO_4 \cdot 2H_2O$.

The majority of wet limestone scrubbing systems use forced oxidation to convert calcium sulfite to gypsum because gypsum is less corrosive (Miller, et al., 2006). As a result, forced oxidation reduces the wear and tear on FGD systems and lowers FGD waste disposal costs (Miller, et al., 2006).

During 1985 to 2010, data suggests that only 41% of the scrubbers that employ limestone as the sorbent material (thereby making them capable of producing a salable gypsum by-product) are successfully able to market their FGD gypsum. The other 59% dispose of their gypsum by-product as a waste material. The remainder of this section explains the emerging market for FGD gypsum focusing on developments in the wallboard industry. Wallboard production accounts for the lion's share of U.S. gypsum consumption, and as a result trends in the wallboard industry are the main determinant of FGD gypsum marketability. Assuming that the ability to sell FGD gypsum is exogenously determined through market contracts, a theoretical model is then developed that compares the profit maximization decisions of an electricity generating facility with a contract to sell gypsum to the profit maximization decisions without a gypsum contract. The theoretical model focuses on the effects of gypsum sales contracts on scrubber efficiency, sorbent use, coal consumption, and coal quality as measured by sulfur content. These parameters in turn determine SO_2 emissions, which is the focus of the empirical section of the paper.

2.1 Wallboard Market

In the U.S., wallboard manufacturers account for roughly 90% of domestic gypsum consumption (USGS, 2012a). Traditionally, U.S. wallboard manufacturing plants have adopted a vertically integrated structure whereby the wallboard firm owns both the gypsum mine and the wallboard manufacturing facility. Furthermore, the manufacturing facilities are generally located in direct proximity of the mine or port of entry in order to minimize the transportation costs of gypsum (O'Brien et al., 1984). As the market for FGD gypsum has developed, wallboard manufacturers using synthetic gypsum have followed similar patterns as their counterparts using mined gypsum and located directly next to FGD gypsum producing power plants in order to minimize gypsum transportation costs. Table 1 summarizes the number of FGD gypsum sellers and crude gypsum mines by census region and district. The most striking pattern revealed in Table 1 is that wallboard manufacturers employing synthetic gypsum in their production processes are generally located in regions with insufficient sources of naturally mined gypsum. Intuitively, FGD gypsum may serve as a cost effective input alternative in regions without mined gypsum sources due to advantages in transportation costs. Transportation costs alone, however, are not sufficient to explain the distribution of the wallboard manufacturing facilities reported in Table 1. The three primary factors explaining the location choices of wallboard manufacturing plants capable of utilizing FGD gypsum are given by the following: (1) availability of high sulfur coal, (2) advantages in transportation costs, and (3) additional capital requirements for processing FGD gypsum.

The availability of high sulfur coal helps explain the lack of wallboard manufacturers using FGD gypsum in the New England division despite an absence of gypsum mines. In

New England, natural gas is the dominant fuel used for electricity generation, and coal accounts for only 2% of total electricity generation (EIA, 2012). As a result, there are few viable FGD gypsum sources available in New England.

On the other hand, the availability of coal with higher sulfur content and lack of a mined gypsum source also helps explain the preponderance of wallboard manufacturers in the South Atlantic, East North Central, and East South Central U.S. census divisions that use FGD gypsum as their raw material. According to the EIA (2012) and USGS (2012b), these three divisions account for roughly 78% of the total bituminous coal consumption, 68% of the total sulfur burned at coal fired electricity generating plants, and only 16% of the total U.S. mined gypsum.[75] The use of high sulfur coal and absence of gypsum mines results in a large FGD gypsum supply in these divisions with lower transportation costs in comparison to mined gypsum alternatives.

One unexplained feature of the wallboard market, however, is that existing wallboard manufacturing plants using mined gypsum as their input buy little if any FGD gypsum from neighboring electricity generating plants. There are two primary explanations for this feature of the wallboard market. First, as Table 1 illustrates, the majority of gypsum mines are located in the Western U.S. region. Coincidentally, the electricity plants in this region generally use the low-sulfur subbituminous coal prevalent in the West (EIA, 2012). As a result, many of these electric plants do not install scrubbers, and the ones that do choose to

---

[75] For disclosure reasons, the USGS does not release statistics by census division. Data are released by clusters of states that may overlap multiple divisions. In order to estimate the total crude gypsum mined by census division, it was assumed that mining totals were distributed evenly over states within clusters.

install scrubbers generally produce a lower quantity of the gypsum by-product than what would be required to support a dedicated wallboard plant using FGD gypsum.

The second primary explanation for wallboard companies being dedicated primarily to the use of either mined or FGD gypsum is that the technology required to manufacture wallboard from the two alternative gypsum sources is not completely interchangeable. As indicated by Srivastava (2000, p. 9) "Since most existing wallboard plants in the United States were designed to use mined rock gypsum as feed material, the solids handling equipment at these plants can use only a limited quantity of FGD gypsum, which has different handling properties."

In sum, wallboard plants capable of using FGD gypsum are more likely to enter markets that are further away from the incumbent firms using mined gypsum that are primarily located in the Western U.S. Furthermore, existing wallboard plants designed to use mined gypsum are not likely to buy large amounts of FGD gypsum due to the additional capital requirements necessary for processing FGD gypsum and the lack of nearby FGD gypsum sources producing substantial quantities of the by-product. Given these features of the wallboard market it is possible to analyze the profit maximization decisions of a scrubber operating electricity plant when the entrance of an FGD gypsum capable wallboard manufacturing facility makes it possible to negotiate gypsum sales contracts.

2.2 Electric Utility Profit Maximization

For simplicity, in the following theoretical model, it is assumed that coal quality is allowed to vary in a continuous fashion with respect to sulfur content, but the heating value and other quality characteristics of coal are held constant. A given coal-fired electricity plant is assumed to operate in one of two mutually exclusive states. In the first state, the electric plant has a contract to sell FGD gypsum and the resulting profits are given by the following:

$$(3). \qquad \pi^{C} = \max_{su,co,ls,ef} \left\{ \begin{array}{c} P_E{*}E(co) + NP_G{*}G(su,co,ls,ef) - P_{co}(su){*}co \\ - P_{ls}{*}ls - P_{SO_2}{*}(SO_2(su,co,ls,ef) - A_0) - C_{ef}(ef) \end{array} \right\},$$

where the quantity of coal, *co*, sulfur content of the coal, *su*, quantity of limestone sorbent material, *ls*, and scrubber operating efficiency, *ef*, are chosen to maximize profits, π^C, in the presence of a gypsum contract. In equation (3), $E(\cdot)$ is the quantity of electricity generated, $G(\cdot)$ is the production function for FGD gypsum, $SO_2(\cdot)$ is the quantity of SO_2 emitted by the plant, and A_0 is the initial quantity of SO_2 allowances distributed to the electricity plant. If

$A_0 > SO_2(\cdot)$ the plant is a net seller of SO_2 emission permits, and if $A_0 < SO_2(\cdot)$ the plant is a net buyer of the tradable permits. Prices for electricity, coal, sorbent, and SO_2 emissions are given by P_E, $P_{co}(\cdot)$, P_{ls}, and P_{SO_2} respectively. The net price of FGD gypsum, NP_G, is equal to the sale price of gypsum net of any transportation/processing cost incurred delivering gypsum to the market. Finally, the cost of increasing scrubber operating efficiency is given by C_{ef}.

In the second possible operating state, the electric plant does not have a contract to sell its synthetic gypsum by-product and profits are given by:

$$(4). \qquad \pi^{NC} = \max_{su,co,ls,ef} \left\{ \begin{array}{c} P_E{*}E(co) - C_G{*}G(su,co,ls,ef) - P_{co}(su){*}co \\ - P_{ls}{*}ls - P_{SO_2}{*}(SO_2(su,co,ls,ef) - A_0) - C_{ef}(ef) \end{array} \right\}.$$

where all the variables are defined as in equation (3), except the term C_G is used to represent the constant per unit disposal cost for the unsold gypsum waste. Assuming there are no costs associated with negotiating gypsum contracts, comparison of equations (3) and (4) reveals that firms will contract to sell their gypsum by-product as long as the net price of gypsum is greater than or equal to the additive inverse of the disposal cost of gypsum ($NP_G \geq -C_G$).

The shape of the production function for gypsum, the emission function for SO_2, and the cost of operating efficiency are characterized by the following monotonicity and curvature conditions:

(5). $$\frac{dG(\cdot)}{d(\cdot)} > 0, \frac{d^2G(\cdot)}{d(\cdot)^2} \leq 0,$$

(6). $$\frac{dSO_2}{dsu} > 0, \frac{d^2SO_2}{dsu^2} \geq 0, \frac{dSO_2}{dco} > 0, \frac{d^2SO_2}{dco^2} \geq 0,$$

$$\frac{dSO_2}{dls} < 0, \frac{d^2SO_2}{dls^2} \geq 0, \frac{dSO_2}{def} < 0, \frac{d^2SO_2}{def^2} \geq 0,$$

(7). $$\frac{dC_{ef}}{def} > 0, \frac{d^2C_{ef}}{def^2} \geq 0.$$

According to (5) the gypsum production function is concave shaped with respect to operating efficiency, sulfur content, coal consumption, and limestone use. Equations (1) and (2) suggests a stoichiometric ratio of 1 mole of gypsum per mole of sorbent or SO_2. Formally, the gypsum production function can be written using the following specifications:

(8). $$G(su,co,ls,ef) = ls * \frac{1\, mole\, G}{1\, mole\, ls} * \frac{mol.\, wt.\, G\, (g/mole)}{mol.\, wt.\, ls\, (g/mole)} * r_{ls}(su,co,ls,ef),$$

(9). $$G(su,co,ls,ef)=ISO_2*\frac{1\,mole\ G}{1\,mole\ SO_2}*\frac{mol.\ wt.\ G\ (g/mole)}{mol.\ wt.\ SO_2\ (g/mole)}*r_{SO_2}(su,co,ls,ef)\,.$$

In equation (8) gypsum production is a function of the quantity of limestone, the constant stoichiometric ratio of gypsum to limestone, $\frac{1\,mole\,G}{1\,mole\,ls}$, the constant molecular weight ratio of gypsum to limestone, $\frac{mol.\ wt.\ G\ (g/mole)}{mol.\ wt.\ ls\ (g/mole)}$, and the reaction rate of limestone, r_{ls}. Likewise, in equation (9) gypsum production is written as a function of the quantity of SO_2 input into the system, ISO_2, the stoichiometric ratio of gypsum to SO_2, the molecular weight ratio of gypsum to SO_2, and the reaction rate of SO_2, r_{SO_2}. The reaction rates r_{ls} and r_{SO_2} measure the fraction of sorbent and SO_2 that actually react in the system. Intuitively increased limestone usage has a direct positive influence on gypsum production, but also lowers the reactivity of the sorbent resulting in diminishing returns. Similarly, increased SO_2 input ($ISO_2=2*su*co$) has a direct positive effect on gypsum production while simultaneously lowering the reaction rate of SO_2.[76]

The SO_2 emission function assumes the following functional form:

(10). $$SO_2(su,co,ls,ef)=ISO_2*(1-r_{SO_2}(su,co,ls,ef))\,.$$

Alternatively, combining equation (10) with equations (8) and (9) results in the following characterization of SO_2 emissions as a function of sorbent usage:

(11). $$SO_2(su,co,ls,ef)=ISO_2-ls*\frac{1mole\ SO_2}{1\,mole\ ls}*\frac{mol.\ wt.\ SO_2\ (g/mole)}{mol.\ wt.\ ls\ (g/mole)}*r_{ls}(su,co,ls,ef)\,,$$

---

[76] SO_2 input is given as 2*su*co* under the assumption that 100% of the total sulfur input (*su*co*) is converted to SO_2. The molecular weight of sulfur is 32 g/mole, and the molecular weight of SO_2 is 64 g/mole. As a result there are 2 grams of SO_2 generated per gram of sulfur assuming 100% conversion.

where all of the variables in equations (10) and (11) are defined the same as in equations (8) and (9). Given the gypsum production functions characterized in equations (8) and (9) and the resulting SO_2 emission functions characterized in equations (10) and (11), it is easily verified that the curvature and monotonicity assumptions on SO_2 emissions given in (6) hold provided that the conditions governing gypsum production given in (5) are satisfied.[77] Furthermore, there are numerous empirically tested FGD chemistry experiments supporting the shape of the SO_2 emission function as characterized by (6) (see, for example, Coutant et al., 1969, Harriott, 1990, Sahar and Kehat, 1991, Maller, 2008, and Brown et al., 2010).

In order to compare the quantity of limestone used in the contract state and the no contract state, the derivatives of the profit functions in both states are evaluated with respect to sorbent usage:

$$(12).\quad \frac{\partial \pi^C}{\partial ls} = NP_G * \frac{\partial G(su,co,ls,ef)}{\partial ls} - P_{ls} - P_{SO_2} * \frac{\partial SO_2(su,co,ls,ef)}{\partial ls} = 0,$$

$$(13).\quad \frac{\partial \pi^{NC}}{\partial ls} = -C_G * \frac{\partial G(su,co,ls,ef)}{\partial ls} - P_{ls} - P_{SO_2} * \frac{\partial SO_2(su,co,ls,ef)}{\partial ls} = 0,$$

Let ls^C and ls^{NC} denote the sorbent quantities satisfying the first order conditions of equations (12) and (13) respectively. Given the monotonicity and curvature conditions specified in equations (5) through (7), it can be shown that $ls^C \geq ls^{NC}$ when $NP_G \geq -C_G$ (i.e., electricity generating plants use more sorbent when they opt to sell their gypsum by-product as opposed to disposing of the by-product as a waste material).[78]

---

[77] Proof is given in appendix proof 1.
[78] Proof is given in appendix proof 2.

Similarly, the differences in average sulfur content, coal consumption, and scrubber efficiency across treatment and control states can be derived by taking the derivative of the profit functions given in equations (3) and (4) with respect to *su*, *co*, and *ef*, respectively. In terms of efficiency, electricity plants increase scrubber efficiency when they choose to sell their FGD gypsum. It is reasonable to assume that the price of coal decreases with respect to sulfur content $\left(\frac{\partial P_{co}}{\partial su} \leq 0\right)$, and electricity generation increases with coal usage $\left(\frac{\partial E}{\partial co} \geq 0\right)$. Unfortunately, the exact nature of the relationship between sulfur content, coal consumption, and gypsum sales cannot be determined without knowing the second order conditions of the hedonic price function for coal, P_{co}, and the electricity production function, *E*. We are unaware of any empirical evidence regarding the shape of the hedonic price function for coal. Empirical evidence regarding the shape of the electricity generation function is mixed, and suggests that *E* may be convex at low levels of coal consumption due to boiler start-up energy losses and concave at higher levels of coal consumption (Tveit et al., 2005; CIAB, 2010). When $NP_G \geq -C_G$, the hedonic price function is convex, and the electricity generation function is concave the following inequalities hold with certainty: $su^C \geq su^{NC}$ and $co^C \geq co^{NC}$. Even if the convexity assumption of P_{co} and the concavity assumption of *E* is violated, it is reasonable to assume the just mentioned sulfur and coal inequalities hold for large quantities of coal use and sulfur content where P_{co} is more likely convex and *E* is likely concave.

In sum, the theoretical model presented suggests that electric plants using SO_2 scrubber technology that have a contract to sell their gypsum by-product will increase their total SO_2 input ($ISO_2 = 2 * su * co$) through increased coal use and increased sulfur content. The theoretical model also suggests that contracting plants will use more sorbent material and increase their operating efficiency. Increased sorbent usage, SO_2 input, and scrubber efficiency all result in increases in gypsum production. However, sorbent use and scrubber efficiency have opposite effects on SO_2 emissions in comparison to SO_2 input. The theoretical model presented in this section is therefore unable to predict the net impact of synthetic gypsum sales on plant-level SO_2 emissions.

According to Srivastava (2000) the actual operating stoichiometry $\left(\frac{moles\ ls}{moles\ ISO_2}\right)$ is a relatively tightly defined parameter for an FGD system. While it is possible that moles of ISO_2 may increase by more/less than the increase in moles of *ls* we expect that any changes in SO_2 emissions from changes in the ratio of these tightly defined parameters will be modest in comparison to the emission reductions resulting from changes in scrubber operating efficiency. As a result, it is expected that SO_2 emissions will decline when plants contract to sell their FGD gypsum by-product. The nature of the relationship between gypsum sales and SO_2 emissions is therefore the empirical question of interest in the remainder of the paper, and the following section describes the data that is available to explore this question.

3. Data

Data on annual power-plant SO_2 emissions comes from the EPA's Air Markets Program Data (AMPD). For each electricity generating facility in the AMPD, the EPA records utility name, facility name, address, a unique longitudinal facility id number, SO_2 emissions, NOx emissions, CO_2 emissions, operating status, primary and secondary fuel type, heat input (a measure of utilization), and any installed abatement technology for SO_2, NOx, and particulate matter. EPA emissions data is available at the boiler-level with a longitudinal facility-boiler id number that allows individual boiler operations within a facility to be tracked over time from 1980 until 2010. The EPA data is available in 5-year intervals from 1980 to 1995, and then collected annually beginning in 1995 corresponding to the start of the U.S. Clean Air Act Amendment (CAAA) emission permit trading program (CAAA, 1990). Because regulated electricity plants are required to hold enough allowances to cover all annual SO_2 emissions it is critical to the success of the CAAA for the EPA to collect precise SO_2 emissions data. To that end, electricity plants install continuous emissions monitoring systems that measure SO_2 flow rates every 15 minutes or less (CEM, 1993). Due to the fact that synthetic gypsum is a by-product produced from scrubbing coal combustion emissions, only boilers using coal as the primary fuel are used for the empirical analysis.

The AMPD data is combined with data on by-product generation, coal quality, and abatement technology from the U.S. Energy Information Administration (EIA).[79] The EIA data is available for larger utilities with 50 or more megawatts generating capacity. In terms of by-product generation, the EIA collects detailed data on synthetic gypsum generation and sales at the facility level on an annual basis. Boiler-level data on coal quality includes average sulfur, ash, and Btu (heat) content. For boilers that install SO_2 scrubbers, the EIA collects data on total tons of sorbent used during the scrubbing process. The two most common types of sorbent used by electric utilities are lime (18% in 2010) and limestone (61% in 2010), both of which generate synthetic gypsum as a by-product. Following the

---

[79] All EPA emission data are available at http://ampd.epa.gov/ampd/ (last accessed October, 2012). The EIA datasets are available at http://205.254.135.7/cneaf/electricity/page/data.html (last accessed October, 2012).

theoretical model presented in section 2, we further restrict our data to only contain boilers with scrubbers installed that use limestone as the sorbent material. Lime scrubbers are omitted from the analysis because lime is more chemically reactive than limestone and lime scrubbers account for only 9% of the observations in the dataset that contract to sell gypsum.

The EIA and EPA data both collect information regarding the installment of pollution abatement technology at the boiler level. Importantly, this allows cross validation between the EPA and EIA data, and results in 36 observations (1.9%) being dropped from the analysis for which the EPA and EIA data are in disagreement regarding installed technology. Data from the EIA is available for all variables beginning in 1985, so the combined panel dataset is an unbalanced panel of electric facility boiler observations from 1985, 1990, and annual observations from 1995 to 2010. There are 238 (12.7%) observations in the data where multiple boilers share the same scrubber. In order to avoid possible bias from aggregating these boilers to the scrubber level, we drop observations of multiple boilers matched to a single scrubber.[80] Table 2 reports variable definitions for the boiler-level data available from the EIA and EPA data sources.

On average, there are 90 scrubber operating boilers available for analysis each year during the twenty-six-year panel. Table 3 presents summary statistics for the boilers over time. Over the entire sample, average yearly SO_2 emissions are declining by roughly 40% between the time periods 1985-1999 and 2000-2010. Total heat input is increasing over these time periods, however suggesting that on average boilers are being utilized more intensively over time.

---

[80] The results are robust to dropping observations of multiple boilers that share a single scrubber or aggregating these boilers to the scrubber level. The results are also robust to the inclusion/exclusion of lime scrubbers.

The average gypsum sold increases by roughly 368% between the time periods 1985-1999 and 2000-2010. This corresponds with a large overall expansion in the synthetic gypsum market over the course of the 26-year panel. Figure 2 presents total gypsum sales for both synthetic and mined gypsum from 1985 to 2010. From 1985 to 1993, synthetic gypsum sales are relatively constant averaging 700 metric tons annually. During the 1985 to 1993 time period synthetic gypsum is primarily produced in the petroleum and chemical industries as a by-product from processes such as acid neutralization. These forms of synthetic gypsum are primarily used in agricultural applications (see, for example, USBM, 1987, and USBM, 1993). Post 1993 there is a significant expansion in the synthetic gypsum market primarily due to increasing sales of FGD gypsum, and by 1996 FGD gypsum accounts for the majority of synthetic gypsum sales (USGS, 1996). Unlike the earlier forms of synthetic gypsum that were produced primarily in the petroleum and chemical industries, FGD gypsum is produced exclusively as a by-product of the SO_2 scrubbing process during electricity generation. As a result of the differing production mechanisms, FGD gypsum is suitable for uses other than agricultural applications, and is primarily sold to the wallboard manufacturing industry (USGS, 1996). Figure 3 reveals a fairly consistent increase in the market share for synthetic gypsum over the same time period between 1996 and 2010. Synthetic gypsum's market share reaches a high of 55% in 2010, but as Figure 2 reveals, the expansion in market share after the 2007 financial crisis is primarily driven by declines in natural gypsum mining as opposed to increases in synthetic gypsum production.

Columns two and three of Table 3 reveal similar trends in SO_2 emissions for boilers that sell gypsum at some point in the 26-year panel in comparison to those that never sell their gypsum by-product suggesting that tightening restrictions of the U.S. CAAA's emission trading program are an effective means of lowering SO_2 emission levels. A naïve difference-in-difference (DID) estimate of the effects of marketing gypsum on SO_2 emissions can be calculated directly from the data in Table 3. SO_2 emissions decline by 5,212 tons from panel A to panel B for gypsum sellers, and only decline by 3,025 tons for non-sellers over the same timeframe. The naïve DID estimate of the effect of FGD gypsum contracts on SO_2 emissions is therefore equal to a 2,187 ton reduction in emissions. It is also worth noting that on average gypsum sellers purchase coal with much larger sulfur content than their non-seller counterparts in both the 1985-1999 and 2000-2010 time periods. On average, gypsum sellers have between 41% and 43% higher sulfur input in comparison to non-sellers. As a final note, gypsum sellers use more sorbent material during the scrubbing process than non-sellers and the average amount of sorbent used is increasing overtime for both sellers and non-sellers.

Overall, the trends revealed in table 3 suggest that electricity plants with contracts to sell their FGD gypsum by-product do indeed use more sorbent material and purchase coal with higher sulfur content as predicted by the theoretical model given in section 2. The following section presents formal tests of the effects of gypsum sales on total boiler-level SO_2 emissions both with and without controls for changes in sorbent use and sulfur input.

4. Results

The key feature of the gypsum market that is exploited in this analysis is the exogenous variation in gypsum sales contracts overtime resulting from the vast expansion of the FGD gypsum market from 1985 to 2010. In 1985, only one U.S. wallboard manufacturing facility was capable of using some FGD gypsum in the production of wallboard. As of 2010, over 30 additional wallboard manufacturing plants in the U.S. and Canada have been built or retrofitted for the purpose of manufacturing wallboard from FGD gypsum.

Given this feature of the data it is possible to analyze the effect of gypsum sales contracts on SO_2 emissions using three primary applications. First, a baseline boiler fixed-effects model is estimated to analyze the net impact of gypsum sales contracts on boiler-level SO_2 emissions. The second application then attempts to isolate the effects of investments in scrubber efficiency by including controls for sorbent use and sulfur input in the estimating equation for SO_2 emissions. In the theoretical model given in section 2, it was hypothesized that sorbent use and sulfur input would both increase as a result of selling synthetic gypsum. Comparison of the estimated coefficient on the sales contract variable between the baseline model and the second application will therefore allow us to determine whether moles of sulfur input increase by more/less than moles of sorbent. Furthermore, the estimated coefficient on contracts in the model that includes controls for sulfur input and sorbent use allows us to isolate the effects of gypsum sellers' increased scrubber operating efficiency in comparison to their non-selling counterparts. Finally, we consider the possibility of omitted variable bias in our application, and use instrumental variable (IV) estimation techniques to estimate the effect of gypsum contracts on SO_2 emissions for plants that are induced to sell FGD gypsum as a result increased gypsum marketability. Each of these empirical applications is presented in turn below.

4.1 Net SO_2 Emissions as a Function of Gypsum Sales

The net effect of gypsum sales on plant-level SO_2 emissions is estimated using the following specification:

(14). $Ln\,Total\ SO_2\ Emissions_{j,t} = \alpha + BC_{j,t}\beta + \delta_1 * Contract_{j,t} + T_t + B_j + SC_{j,t}\theta + MA_{j,t}\gamma + \varepsilon_{j,t},$

where the natural log of total SO_2 emissions at boiler j and time t is a function of a vector of two time varying boiler characteristics ($BC_{j,t}$): *Ln Total Heat Input*, and *Ln Total Ash*. The net-effect model includes year (T_t) and boiler (B_j) fixed effects, a vector of indicator variables controlling for type of scrubber installed ($SC_{j,t}$), and a vector of indicator variables controlling for the manufacturer of any installed scrubbers ($MA_{j,t}$). $Contract_{j,t}$ is an indicator variable equal to one for boilers operated at plants with contracts to sell their FGD gypsum and equal to zero otherwise. The estimated coefficient δ_1 is therefore the key coefficient of interest because it provides the net effect of gypsum sales contracts on SO_2 emissions. Finally, $\varepsilon_{j,t}$ is a random error component that is clustered at the boiler and year levels to allow for correlation within clusters.

The results from equation (14) are presented in column 1 of table 4. The estimated coefficient on heat input is positive as expected, and indicates that a 1% increase in boiler usage is associated with a 1.1% increase in SO_2 emissions. Ash content is positively correlated with SO_2 emissions, but the effect is statistically insignificant. Finally, the estimated impact of a gypsum sales contract given by the coefficient δ_1 is negative as expected, and is statistically significant at the 1% level. The results indicate that the ability to market FGD gypsum is associated with an 18.8% reduction in SO_2 emissions.

A key underlying assumption of the fixed effects model estimated in equation (14) is that SO_2 emissions for contracting and non-contracting plants follow a common trend in the absence of gypsum contracts. In order to test the common trend assumption we drop observations that occur in the year of contract and all subsequent years, and estimate the following equation:

(15).

$$Ln\,Total\ SO_2\ Emissions_{j,t} = \alpha + BC_{j,t}\beta + \delta_1' * Contract_{j,t+n} + T_t + B_j + SC_{j,t}\theta + MA_{j,t}\gamma + \varepsilon_{j,t}\,.$$

All of the variables in equation (15) are defined as in equation (14) except the natural log of SO_2 emissions at time t is now regressed on contract status of boiler j at time t+n. Contract status measured at time t+n serves as a "pseudo" treatment indicator, and it is expected that δ_1' will be statistically insignificant if the common trend assumption holds.

Panel A in the appendix Table A1 provides the results from estimating equation (15) for n=2, 4, 6, and 8. The estimated coefficient on the pseudo contract variable is relatively small (between -.5% and 33% of the 18.8% reduction in SO_2 emissions estimated in equation (14)) and statistically indistinguishable from zero for all choices of n. Because we do not expect the common trend assumption to hold in both logs and levels, we also test for common trend with a linear specification. The results from the linear specification are given in panel B of Table A1. In the linear models, the estimated coefficients on the pseudo contract variable are also statistically insignificant for all choices of n. The standard errors in the linear specification are relatively smaller than the standard errors in the log-log specification, and as a result statistical insignificance holds at higher significance levels in the log-log models. Therefore, the remainder of this analysis uses the log-log specification as the preferred specification.[81]

4.2 Reduced SO_2 Emissions due to Operating Efficiency

The results estimated in equation (14) suggest that on average gypsum sellers are able to reduce SO_2 emissions in comparison to their non-selling counterparts. In order to narrow down the source of emission reductions we estimate the following model:

(16). $Ln\,Total\ SO_2\ Emissions_{j,t} = \alpha + BC_{j,t}\beta + \delta_1 * Contract_{j,t} + T_t + B_j + SC_{j,t}\theta + MA_{j,t}\gamma + \varepsilon_{j,t}$,

where all of the variables are defined as in equation (14), except the vector of boiler characteristics ($BC_{j,t}$) in equation (16) now includes all of the following: *Ln Total Heat Input, Ln Total Ash, Ln Total Sulfur, Ln Total Sorbent,* and *Ln Hours In Service*. The results from the estimation of equation (16) are presented in column 2 of Table 4. The estimated coefficient δ_1 is negative and statistically significant at the 1% level. The results indicate that a contract to sell gypsum is associated with an 18.3% reduction in SO_2 emissions, ceteris paribus. Overall, the results suggest that the majority of emission reductions are due to increased scrubber operating efficiency at contracting boilers. This finding is consistent with

---

[81] In order to test for sensitivity to choice of functional form, the incremental effects of gypsum contracts on SO_2 emissions were estimated using log-log, linear, Box-Cox, and random effects tobit specifications. The log-log specification is given in equation(14). Linear models include higher-order polynomials to allow for flexible functional form. Box-Cox estimation rejects linear, log-log, log-linear, and inverse functional forms. However, among the more popular functional forms, the Box-Cox model seems to more closely resemble the log-log specification. Finally, the random effects tobit model is based on the structural specification given in equation (10). Following equation (10), $r_{SO_2}(su, co, ls, ef)$ is constructed by calculating $1 - \frac{SO_2(su,co,ls,ef)}{ISO_2}$ and then estimated as a function of contract status using random effects tobit where the reaction rate is bound between zero and one. The first four rows of column one in appendix Table A2 provide results from the tests of alternative functional forms, and overall the incremental effects are fairly robust to specification choice. Column two of Table A2 also finds similar incremental effects across specifications for the models that attempt to isolate the impact of scrubber operating efficiency.

the theoretical model presented in section 2, where it was postulated that the ratio of sulfur input to sorbent use is a relatively tightly defined parameter in scrubber operations.

Marketability as an Instrument for Gypsum Contracts

In the previous analysis it was assumed that gypsum contracts affect SO_2 emissions following the path analysis illustrated in Figure 4. According to Figure 4 gypsum sales contracts influence SO_2 emissions by providing incentives for electricity plants to change the following scrubber operating parameters: sulfur input, sorbent use, and operating efficiency.

Recall from the theoretical model presented in section 2, however, firms choose to sell gypsum provided the net price of gypsum is greater than or equal to the additive inverse of disposal costs ($NP_G \geq -C_G$). There is reason to believe that latent unobserved variables may increase disposal costs for some plants and also have a direct effect on SO_2 emissions through sulfur input, sorbent use, and operating efficiency. In June 2010, the EPA proposed a regulation on coal combustion residual disposal in response to the 2008 disposal pond leakage at the Tennessee Valley Authority's Kingston, TN electricity plant (DCREC, 2010). It is reasonable to assume that increased oversight of gypsum disposal increases the cost of disposal and hence increases the likelihood of contracting gypsum sales. It is also reasonable to assume that power plants disposing of gypsum operate their scrubbers less intensively due to increased operating costs.[82] Within this latent variable framework, the directional chain illustrating the impact of gypsum contracts on plant-level SO_2 emissions is given in Figure 5.

---

[82] The EPA proposed regulation on the disposal of coal combustion residuals occurs at the very end of the timeframe analyzed in our sample. It is possible, however that there are state and/or local regulations on gypsum disposal that are not accounted for in our analysis. Failure to account for these laws may bias our estimated coefficient on gypsum contracts.

In order to avoid bias arising from unobserved increased regulation of FGD gypsum disposal, we estimate the following two-stage least squares regression:

(17). $Contract_{j,t} = \alpha_1 + BC_{j,t}\beta_1 + T_t + B_j + SC_{j,t}\theta_1 + MA_{j,t}\gamma_1 + Z_{j,t}\mu + u_{j,t}$,

(18).

$$Ln\,Total\ SO_2\ Emissions_{j,t} = \alpha_2 + BC_{j,t}\beta_2 + \delta_2 * \widehat{Contract}_{j,t} + T_t + B_j + SC_{j,t}\theta_2 + MA_{j,t}\gamma_2 + \varepsilon_{j,t}.$$

Equation (17) illustrates that in the first stage contract status is regressed on the exogenous boiler characteristics, scrubber type indicator variables, manufacturer indicator variables, time and boiler fixed effects, and an exogenous vector of instrumental variables ($Z_{j,t}$) that are correlated with SO_2 emissions only through their influence on gypsum contracts. The vector of instruments includes three indicator variables controlling for the opening of an FGD gypsum capable wallboard manufacturing facility located within 50 miles of the electricity plant, between 50 and 100 miles of the electricity plant, and between 100 and 250 miles of the electricity plant.[83] The vector of instruments also includes two count variables measuring the number of FGD gypsum capable wallboard plants located between 50 and 100 miles of the electricity plant, and the number of wallboard plants located between 100 and 250 miles of the electricity plant.[84]

Results from the first-stage estimates of contract status are presented in Table A3 of the appendix. The estimated coefficients on the indicator variables for 50 and 100 miles are

---

[83] A 250 mile cutoff was used based on the Obrien et al. (1984) marketing study that found electricity plants could successfully market their FGD gypsum to wallboard plants located within a 250 mile radius. Alternative cutoff thresholds were considered and distances greater than 250 miles were excluded because they tended to violate the strong instrument requirements.

[84] A count variable for the number of wallboard plants located within 50 miles of the electricity plant is excluded because it is perfectly collinear with the 50 mile indicator variable.

statistically significant at the 1% level. The coefficient on the 50 mile dummy variable is positive as expected, and indicates that the arrival of a wallboard manufacturing facility within 50 miles of the electricity plant increases the probability of a gypsum sales contract by roughly 40% in both the net effect and efficiency models. The estimated coefficient on the 100 mile indicator variable has a counterintuitive negative sign, but the negative effect is offset by the positive estimated coefficient on the 100 mile count variable. Finally, the estimated coefficient on the 250 mile indicator variable is positive and statistically insignificant, but the estimated coefficient on the 250 mile count variable is positive and significant at the 10% level. F-statistics are also reported for the first-stage regressions that test the joint null hypothesis that all excluded instruments ($Z_{j,t}$) are statistically insignificant.[85] The F-statistics are both greater than 18 in the net effect and efficiency specifications, and indicate that the joint null hypothesis of statistical insignificance can be rejected at the 99% confidence level. Bound et al. (1995) suggest an F-statistic cutoff threshold of 10 for assessing the predictive strength of IVs, and our instrument set satisfies this strong instrument requirement.

The predicted contract status from the first-stage regressions is then used to estimate the second-stage regression given by equation (18). Results from the second-stage regressions are presented in columns three and four of Table 4 for the net effect and efficiency specifications, respectively. Similar to the net effect and efficiency specifications that do not use IVs (columns one and two of Table 4), the IV results indicate that the

---

[85] The reported F-statistics are *Kleibergen-Paap F-statistics that correct for boiler and year clustered standard errors.*

majority of SO_2 emission reductions can be attributed to increased scrubber operating efficiency. The estimated coefficients on sales contract in the IV models, however, indicate a much larger reduction in SO_2 emissions than what was predicted in the models that do not correct for the omitted variable bias. The net effect IV model suggests that sales contracts are associated with a 39% reduction in SO_2 emissions, and the effect is significant at the 1% significance level. The efficiency IV model estimates a 37.5% reduction in SO_2 emissions following a contract agreement, and the effect is significant at the 5% significance level. The fifth row of Table A2 in the appendix provides the associated incremental effects of the log-log IV specifications from Table 4. For comparison purposes to alternative specifications the incremental effects from a linear IV specification are also reported in the last row of Table A2. The linear IV model estimates a larger reduction in SO_2 emissions, but the results are similar to the preferred log-log IV specification.

Finally, Table 4 presents tests for overidentification, underidentification, and exogeneity of the contract indicator variable. The Davidson-Mackinnon exogeneity tests reject the null hypothesis that contracts are exogenous, and the IV models are therefore chosen as the preferred results. The overidentification test statistics are statistically insignificant at a level greater than 15%, and indicate that the instruments are valid in our application. Finally, the underidentification test statistic is significant at the 10% level and provides further evidence that the set of instruments is relevant.

Overall, the results from equation (18) suggest that synthetic gypsum contracts reduce total SO_2 emissions, but the majority of emission reductions are due to increased scrubber operating efficiency. Srivastava (2000) and Walsh, et al. (2006) identify several additional parameters that can affect scrubber operating efficiency: limestone dissolution rates (determined by slurry pH, and limestone particle size), limestone purity, temperature, and reaction time (determined by the size of the reaction tank). The data used in this analysis do not allow us to identify the exact sources of the increased operating efficiency, but these are interesting questions for future research.

5. Conclusion

This research estimates the impact of boiler-level changes in gypsum sales contracts on SO_2 emissions at coal-fired boilers that operate SO_2 scrubbers. The results suggest that SO_2 emissions decline as a result of electricity generating plants ability to negotiate contracts to sell their FGD gypsum by-product. The reductions in SO_2 emissions are primarily driven by increased scrubber operating efficiency at contracting plants. On average, the net results indicate that a gypsum sales contract results in a 39% reduction in SO_2 emissions holding other factors constant. The results are consistent with the maintained hypothesis that the ability to sell synthetic gypsum by-product creates incentives for power plants to increase the operating efficiency of their SO_2 scrubber systems.

Estimated benefits of SO_2 emission reductions range anywhere from $2,400 per ton to $112,000 dollars per ton (see, for example, Banzhaf et al., 2004, Fann et al., 2009, EPA, 2010, and Shadbegian et al., 2006).[86] The lower bound benefit estimate of $2,400 per ton of reduced SO_2 emissions comes from Banzhaf et al. (2004), and is calculated using a $3.0 million measurement of the value of a statistical life (VSL) based on a VSL meta-analysis conducted by Mrozek and Taylor (2002). The upper bound benefit estimate of $112,000 per

---

[86] All $ values are measured in 2012 dollars unless otherwise noted.

ton comes from the U.S. EPA (2010), and is calculated using a central point estimate of $8.2 million for the VSL.[87] The U.S. EPA's (2010) lower bound estimate is $47,000 per ton of SO_2 abatement. To date, the EPA has conducted the most comprehensive evaluations of the costs and benefits of the Clean Air Act, and for the remainder of this analysis we estimate the range of benefits associated with synthetic gypsum production using the EPA's upper and lower bound benefit per ton estimates.

There are a total of 131 boilers distributed across 63 electricity plants in our data set that sell synthetic gypsum at some point during the time period 1985 to 2010. The net impact of a gypsum sales contract is estimated to be a 3,494 ton reduction in total annual boiler SO_2 emissions. The following formula is used to calculate the total annual benefits associated with reductions in SO_2 emissions as a result of synthetic gypsum sales:

(19). $Total\ Benefits = 3{,}494 * Total\ Number\ of\ Sellers * Benefits\ Per\ Ton\ SO_2\ Reduction$.

Table 5 presents the yearly total benefit estimates associated with reduced SO_2 emissions from power plants selling synthetic gypsum measured in billions of dollars. On average across all years in the panel, gypsum sales generate between $7.0 and $16.7 billion dollars in total benefits largely resultant from reduced particulate matter concentrations and their associated reductions in individual mortality risks. The last two rows of Table 5 summarize the total benefits arising from gypsum inspired reductions in SO_2 emissions. The total benefits from 1990 to 2010 associated with synthetic gypsum production are between $147.1

---

[87] Banzhaf et al. (2004) also use the lower bound epidemiological estimates of the effects of particulate matter on mortality risks that were calculated by Pope et al. (2002). The U.S. EPA's $112,000 per ton benefit estimate is based on an epidemiology study by Laden et al. (2006), which estimates a much more severe impact of particulate matter on mortality risk. Based on the Pope et al. (2002) epidemiology study, the U.S. EPA's estimate of benefit per ton of SO_2 abatement is $47,000.

billion (PV = $73.5 billion) and $350.6 billion (PV = $152.4 billion). Overall, the results indicate significant benefits associated with the sales of FGD by-product gypsum. At first glance, the magnitude of these benefits may appear extreme. In comparison, the U.S. EPA (2011) estimates the total benefits of the CAA from 1990 to 2010 to be approximately $16.6 trillion (PV = $8.6 trillion). Our results indicate that between 0.9% and 2.1% of the total benefits of the CAA between 1990 and 2010 can be attributed to the emergence of a marketable synthetic gypsum by-product.[88]

The benefit analysis presented above highlights the importance of understanding the methods employed by electric generating utilities when complying with the standards established by the CAAA. The ability to sell FGD gypsum creates incentives for plants to operate their SO_2 scrubbers more efficiently, and results in significant external benefits largely in the form of reductions in morbidity and mortality risk. Furthermore, Carlson et al. (2000) find that the EPA's original compliance cost estimates of the CAAA were overstated because the EPA did not account for declining prices of low sulfur coal and advances in fuel switching technology. Presumably, plants choosing to sell their FGD gypsum are also able to achieve cost savings, which offers an additional explanation of the overestimated compliance costs originally estimated by the EPA.

---

[88] When calculating the benefits as a percentage of the total CAAA benefits, we do not account for general equilibrium effects in terms of allowance prices or rebound effects from utilities increasing emissions at other units or plants.

REFERENCES

Banzhaf, S., D. Burtraw and K. Palmer (2004): "Efficient Emission Fees in the U.S. Electricity Sector," *Resource and Energy Economics*, 26, 3, 317-341.

Bound, J., D. Jaeger and R. Baker (1995): "Problems with Instrumental Variables Estimation when the Correlation between the Instruments and Endogenous Explanatory Variable is Weak," *Journal of the American Statistical Association*, 90, 430, 443-450.

Brown, S., R. DeVault and P. Williams (2010): "Determination of Wet FGD Limestone Reactivity," *Technical Paper: Presented to Electric Power*, 1-9.

Carlson, C., D. Burtraw, M. Cropper and K. Palmer (2000): "Sulfur Dioxide Control by Electric Utilities: What are the Gains from Trade?" *Journal of Political Economy*, 108, 6, 1292-1326.

U.S. Congress (1990) Clean Air Act Amendments, < http://epa.gov/ttn/oarpg/gen/caa-pdf.pdf>.

Environmental Protection Agency (2005) Clean Air Interstate Rule, *Federal Register*, 70, 91, 25162-25405.

Coal Industry Advisory Board (2010): "Power Generation from Coal: Measuring and Reporting Efficiency Performance and CO_2 Emissions," Paris: International Energy Agency.

US Government Printing Office (1993) Continuous Emission Monitoring, Title 40: Protection of Environment [40 CFR], part 75.

Coutant, R., R. Barrett, R. Simon, B. Campbell and E. Lougher (1969): "Investigation of the Reactivity of Limestone and Dolomite for Capturing SO2 from Flue Gas," Columbus: National Air Pollution Control Administration.

Environmental Protection Agency (2010) Disposal of Coal Combustion Residuals from Electric Utilities, *Federal Register*, 75, 118, 35127-35264.

Duan, N. (1983): "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, 78, 383, 605-610.

Fann, N., C. Fulcher and B. Hubbell (2009): "The Influence of Location, Source, and Emission Type in Estimates of the Human Health Benefits of Reducing a Ton of Air Pollution," *Air Quality, Atmosphere & Health*, 2, 3, 169–176.

Flagan, R. and J. Seinfeld (1988), *Fundamentals of Air Pollution Engineering*, Prentice Hall, Englewood Cliffs: Dover Publications.

Harriott, P. (1990): "A Simple Model for SO_2 Removal in the Duct Injection Process," *Journal of Air Waste Management Association*, 40, 7, 998-1003.

Henzel, D., B. Laseke, E. Smith and D. Swenson (1981): "Limestone FGD Scrubbers: User's Handbook," Research Triangle Park: U.S. Environmental Protection Agency, Office of Research and Development.

Laden, F., J. Schwartz, F. Speizer, and D. Dockery (2006): "Reduction in Fine Particulate Air Pollution and Mortality," *American Journal of Respiratory and Critical Care Medicine*, 173, 6, 667-672.

Maller, G. (2008): "Wet FGD Chemistry and Performance Factors," *Wet FGD Technical Seminar*, Worldwide Pollution Control Association, Orlando, 1-68.

Miller, C., T. Feeley, III, W. Aljoe, B. Lani, K. Schroeder, C. Kairies, A. McNemar, A. Jones and J. Murphy (2006): "Mercury Capture and Fate Using Wet FGD at Coal-Fired Power Plants," *DOE/NETL mercury and wet FGD R&D*.
<http://www.netl.doe.gov/technologies/coalpower/ewr/coal_utilization_byproducts/pdf/mercury_%20FGD%20white%20paper%20Final.pdf >

Mrozek, J. and L. Taylor (2002): "What Determines the Value of a Statistical Life? A Meta-Analysis," *Journal of Policy Analysis and Management*, 21, 2, 253-270.

O'Brien, W., W. Anders, R. Dotson, J. Veitch and J. Jones (1984): "Marketing of By-product Gypsum from Flue Gas Desulfurization," Washington, DC: U.S. Environmental Protection Agency, Office of Research and Development.

Ozkaynak, H. and J. Spengler (1985): "Analysis of Health Effects Resulting from Population Exposures to Acid Precipitation Precursors," *Environmental Health Perspectives*, 63, 45-55.

Pope, C. III., R. Burnett, M. Thun, E. Calle, D. Krewski, K. Ito and G. Thurston (2002): "Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution," *Journal of the American Medical Association*, 287, 9, 1132-1141.

Sahar, A. and E. Kehat (1991): "Sulfur Dioxide Removal from Hot Flue Gases by Lime Suspension Spray in a Tube Reactor," *Industrial & Engineering Chemistry Research*, 30, 3, 435-440.

Shadbegian, R., W. Gray and C. Morgan (2006): "Benefits and Costs from Sulfur Dioxide Trading: A Distributional Analysis," In Gerald R. Visgilio and Diana M. Whitelaw, eds. *Acid in the Environment: Lessons Learned and Future Prospects*, pp. 241-259. New London: Springer US.

Srivastava, R. (2000): "Controlling SO_2 Emissions: A Review of Technologies," *Research Triangle Park: US Environmental Protection Agency National Risk Management Research Laboratory*.

Tveit, T-M, T. Savola and C-J. Fogelholm (2005): "Modelling of Steam Turbines for Mixed Integer Nonlinear Programming (MINLP) in Design and Off-Design Conditions of CHP Plants," In J. Amudsen, H.I. Andersson, E. Celledoni, T. Gravdahl, F.A. Michelsen, H.R. Nagel, T. Natvig, eds. *Proceedings of the SIMS2005 – 46th Conference on Simulation and Modeling*, pp. 335–344. Trondheim, Norway,

US Bureau of Mines (1987) Minerals Yearbook 1985, Washington, DC: U.S. Department of the Interior, U.S. Bureau of Mines.

US Bureau of Mines (1993) Minerals Yearbook 1990, Washington, DC: U.S. Department of the Interior, U.S. Bureau of Mines.

US Energy Information Administration (2012) Electric Power Monthly with Data for June 2012, Washington, DC: U.S. Department of Energy.

US Environmental Protection Agency (2010): "Final Regulatory Impact Analysis (RIA) for the SO_2 National Ambient Air Quality Standards (NAAQS)," Office of Air Quality Planning and Standards, Research Triangle Park: U.S. Environmental Protection Agency.

US Environmental Protection Agency (2011): "The Benefits and Costs of the Clean Air Act from 1990 to 2020," Office of Air and Radiation, Washington, DC: U.S. Environmental Protection Agency.

US Geological Survey (1996) Minerals Yearbook 1996, Reston: U.S. Department of Interior, U.S. Geological Survey.

US Geological Survey (2012a) Mineral Commodity Summaries 2012, Reston: U.S. Department of Interior, U.S. Geological Survey.

US Geological Survey (2012b) Minerals Yearbook 2010, Reston: U.S. Department of Interior, U.S. Geological Survey.

Walsh, M., M. Mengel, A. Evans, E. Gal, G. Cavallari, M. Bienati and P. Cavezzale (2006): "Parameters Impacting Limestone Dissolution in Flue Gas Desulfurization (FGD) Systems," Energia da Carbone Pulito nella Prospettiva del dopo Kyoto, Associazione Termotecnica Italiana, Milan, pp. 1-9.

Table III-1: Location of FGD Gypsum Producers and Gypsum Mines

Census Region / Division	# FGD Gypsum Producers[a]	#Mines[b]
North East		
New England	0	0
Middle Atlantic	7	0
South		
South Atlantic	23	1
East South Central	9	1
West South Central	4	11
Midwest		
East North Central	17	7
West North Central	2	11
West		
Mountain	0	17
Pacific	1	6
Total:	63	54

[a] Data on gypsum producers is available from the EIA electricity data collected on form 923.
[b] Data on crude gypsum mines is available from USGS (2012b).

Table III-2: Variable Description for EIA and EPA Data

Variable Name	Description
Total SO_2 Emissions	Total annual SO_2 emissions measured in tons.
Total Heat	Total annual heat input measured in trillions of British thermal units (Btu).
Total Sulfur	Total sulfur input measured in tons
Total Ash	Total ash input measured in tons
Total Sorbent	Total annual limestone sorbent usage measured in tons.
Total Gypsum Sold	Total annual gypsum by-product sales measured tons.
Hours In Service	Total annual scrubber operating hours.
Contract	Indicator variable equal to one if plants have contracts to sell FGD gypsum.

Table III-3: Average Annual Summary Statistics [a]

Time Period	Total Sample Mean (Std. Dev.)	Gypsum Contract Mean (Std. Dev.)	No Gypsum Contract Mean (Std. Dev.)
		Panel A: 1985-1999	
Total SO_2 Emissions	9.978	12.10	8.002
	(9.556)	(10.80)	(7.745)
Total Heat	0.0333	0.0348	0.0319
	(0.0192)	(0.0216)	(0.0166)
Total Sulfur	23.90	30.28	17.95
	(20.54)	(21.47)	(17.71)
Total Ash	180.1	183.3	177.1
	(184.0)	(186.7)	(181.8)
Total Sorbent	73.30	97.64	50.65
	(69.68)	(72.67)	(58.39)
Total Gypsum Sold	8.487	17.61	-----
	(31.44)	(43.51)	-----
Hours In Service	7.260	7.387	7.142
	(1.385)	(1.022)	(1.646)
# obs.	529	255	274
		Panel B: 2000-2010	
Total SO_2 Emissions	6.025	6.888	4.977
	(6.070)	(6.562)	(5.232)
Total Heat	0.0365	0.0362	0.0368
	(0.0185)	(0.0190)	(0.0177)
Total Sulfur	26.59	32.94	18.88
	(21.23)	(21.61)	(17.98)
Total Ash	189.6	184.3	196.1
	(168.4)	(169.8)	(166.8)
Total Sorbent	90.06	111.4	64.19
	(74.66)	(74.17)	(66.73)
Total Gypsum Sold	39.75	72.49	-----
	(89.72)	(111.0)	-----
Hours In Service	7.599	7.602	7.596
	(1.040)	(0.921)	(1.169)
# obs.	1,001	549	452

[a] All data reported in thousands unless otherwise noted. All EPA data are from the EPA AMPD data, available as downloadable files from: http://ampd.epa.gov/ampd/ (last accessed October, 2012). The EIA data are available as downloadable files from: http://205.254.135.7/cneaf/electricity/page/data.html (last accessed October, 2012).

Table III-4: Estimates of Plant-level Changes in Natural Log SO_2 Emissions [a]

	Estimated Coefficients (Std. Errors)			
Variable Name	**Net Effect**	**Efficiency**	**Net Effect - IV**	**Efficiency - IV**
Ln Total Heat	1.055***	1.162***	0.986***	1.093***
	(0.091)	(0.101)	(0.102)	(0.109)
Ln Total Ash	0.036	-0.090	0.065	-0.062
	(0.068)	(0.104)	(0.074)	(0.106)
Ln Total Sulfur	-----	0.194*	-----	0.191*
		(0.106)		(0.109)
Ln Total Sorbent	-----	-0.053*	-----	-0.056*
		(0.031)		(0.032)
Ln Hours In Service	-----	-0.185***	-----	-0.174**
		(0.068)		(0.069)
Contract	-0.208***	-0.202***	-0.495***	-0.470**
	(0.073)	(0.069)	(0.189)	(0.188)
Number of obs.	1,530	1,527	1,530	1,527
R-squared	0.472	0.488	0.454	0.472
Number of boilers	153	153	153	153
F-Statistic	-----	-----	18.780	18.623
F-Stat. p-value	-----	-----	(0.000)	(0.000)
Overidentification	-----	-----	6.603	6.416
Overid. p-value	-----	-----	(0.158)	(0.170)
Underidentification	-----	-----	9.469	9.512
Underid. p-value	-----	-----	(0.092)	(0.090)
Davidson-Mackinnon			10.760	9.596
David.-Mack. p-value			(0.001)	(0.002)

[a] Statistical significance at the 1%, 5%, and 10% level are represented by ***,**,and *, respectively. Although not reported, the models presented also include a full set of boiler and year fixed effects, scrubber type dummies, and scrubber manufacturer dummies.

Table III-5: Total Benefits of Reduced SO_2 Emissions Resulting from FGD Gypsum Sales [a]

Year	Total Annual Benefits – \$47,000 per ton SO_2 (billions\$)[b]	Total Annual Benefits – \$112,000 per ton SO_2 (billions\$)[c]
1990	0.985	2.348
1991	1.248	2.974
1992	1.511	3.6
1993	1.774	4.226
1994	2.036	4.853
1995	2.299	5.479
1996	3.12	7.435
1997	3.941	9.392
1998	4.434	10.566
1999	5.584	13.306
2000	6.569	15.654
2001	7.226	17.219
2002	7.883	18.784
2003	8.375	19.958
2004	8.375	19.958
2005	8.54	20.35
2006	9.689	23.089
2007	10.839	25.828
2008	15.765	37.569
2009	17.736	42.265
2010	19.214	45.787
Total:	147.143	350.64
PV (5% Discount Rate)	73.508	175.168

[a] The data used in this analysis does not include observations for years 1991 to 1994 and 2006. Benefits for the missing years are linearly extrapolated.

[b] The lower bound of \$47,000 per ton of SO_2 abatement is based on the study of epidemiological impacts conducted by Pope et al. (2002).

[c] The upper bound is based on the epidemiological study by Laden et al. (2006).

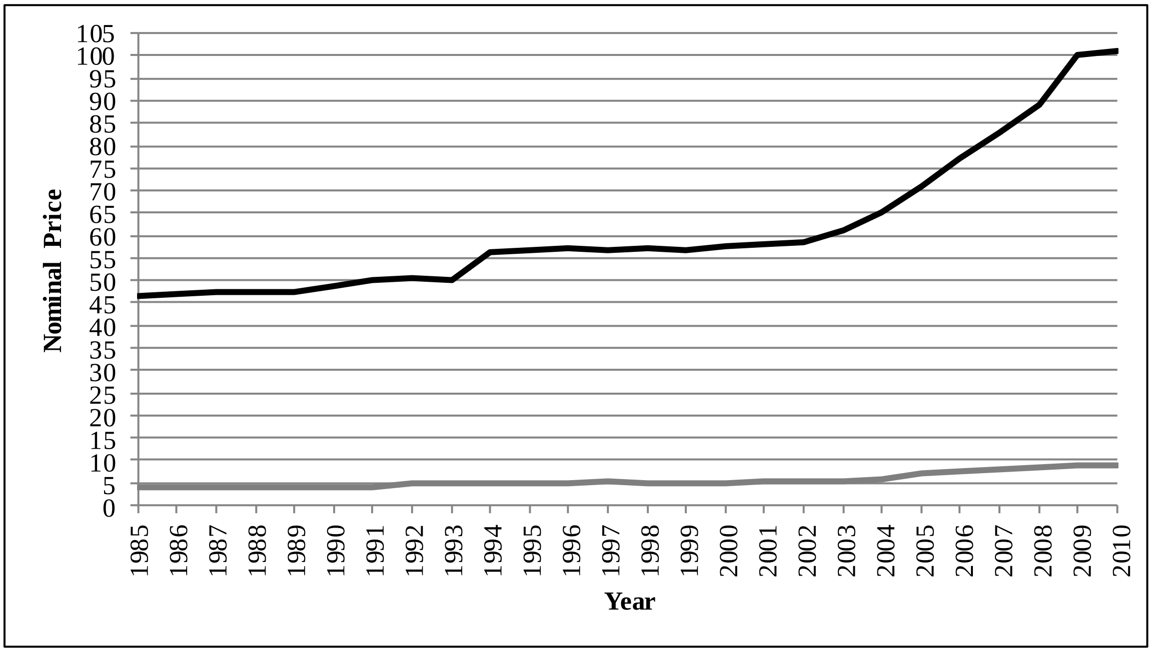

Figure III-1: Lime and Limestone Average Price per Metric Ton (1985-2010)

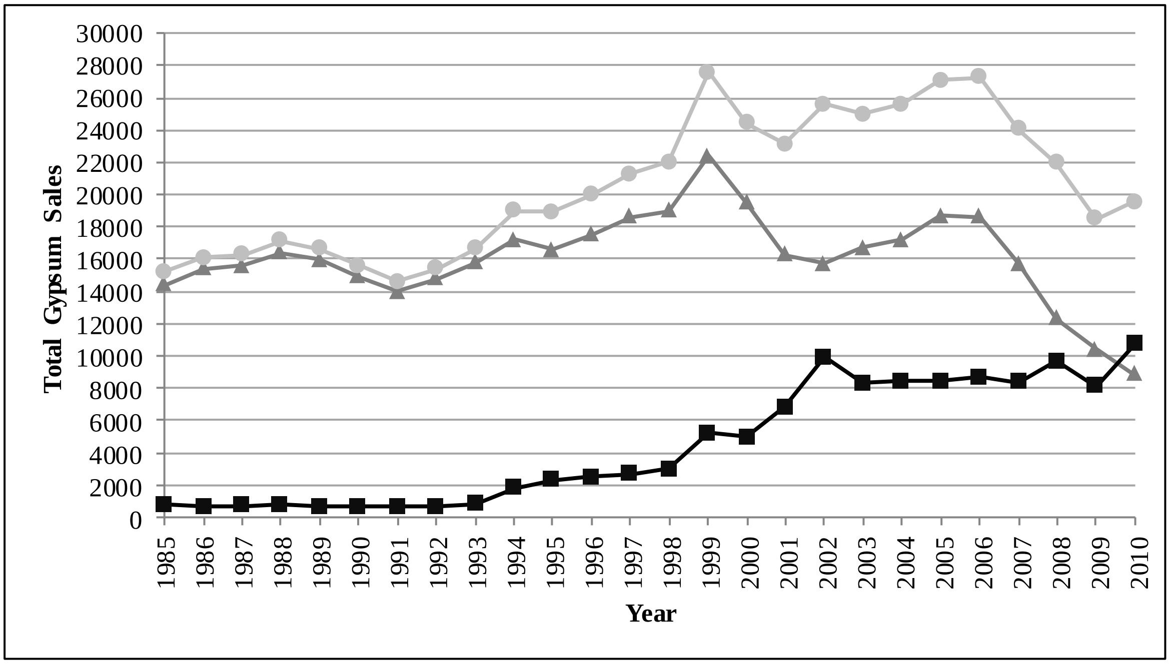

Figure III-2: Gypsum Sales 1985-2010 (Measured in Thousand Metric Tons)

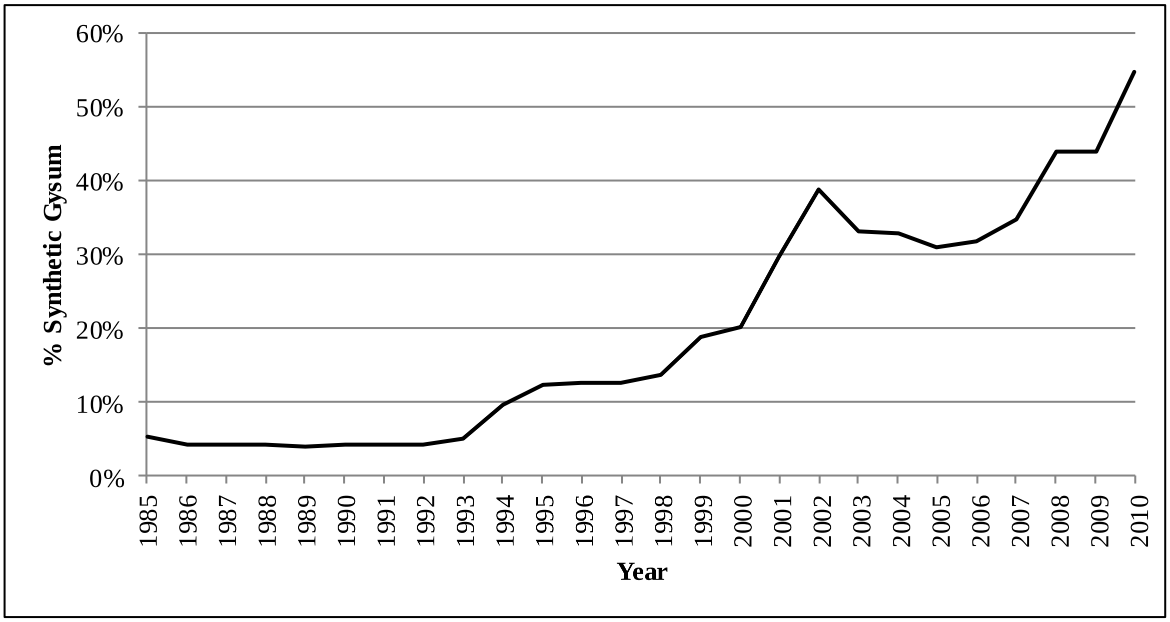

Figure III-3: Synthetic Gypsum Market Share

Figure III-4: Path Analysis Diagram for Gypsum Sales Contracts and SO2 Emissions

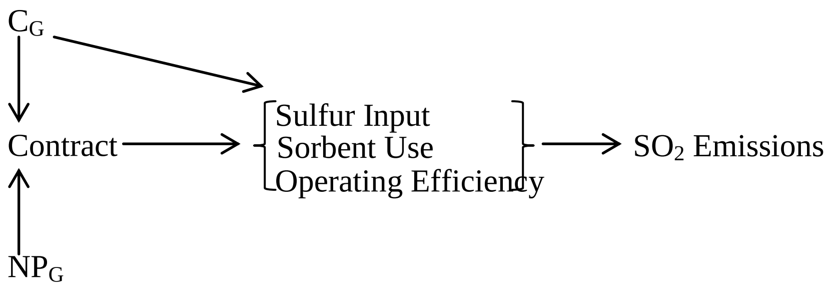

Figure III-5: Path Analysis Diagram with Unobserved Disposal Regulations

APPENDICES

Appendix A

Proof 1:

Gypsum production is characterized by the following:

(A-1). $G(su, co, ls, ef) = ls * \alpha * r_{ls}(su, co, ls, ef)$,

where a is a constant term equal to $\frac{1\, mole\, G}{1\, mole\, ls} * \frac{mol.\ wt.\ G\ (g/mole)}{mol.\ wt.\ ls\ (g/mole)}$, which is a positive valued expression. Differentiating (A-1) with respect to ls yields the following:

(A-2). $$\frac{dG(su, co, ls, ef)}{dls} = \alpha * r_{ls}(su, co, ls, ef) + ls * \alpha * \frac{dr_{ls}(su, co, ls, ef)}{dls},$$

Assuming the derivative of the reaction rate of limestone with respect to limestone is negative, the gypsum production function will be positively sloped in terms of limestone use provided that the following inequality holds:

(A-3). $$\alpha * r_{ls}(su, co, ls, ef) > -ls * \alpha * \frac{dr_{ls}(su, co, ls, ef)}{dls}.$$

The SO_2 emission function is defined as the following:

(A-4). $SO_2(su, co, ls, ef) = ISO_2 - ls * b * r_{ls}(su, co, ls, ef)$,

where b is equal to the positive constant $\frac{1\, mole\, SO_2}{1\, mole\, ls} * \frac{mol.\ wt.\ SO_2\ (g/mole)}{mol.\ wt.\ ls\ (g/mole)}$. Differentiating equation (A-4) with respect to ls yields the following:

(A-5). $$\frac{dSO_2(su, co, ls, ef)}{dls} = -b * r_{ls}(su, co, ls, ef) - ls * b * \frac{dr_{ls}(su, co, ls, ef)}{dls},$$

If (A-3) is satisfied the following inequality also holds:

(A-6). $$-b * r_{ls}(su, co, ls, ef) < ls * b * \frac{dr_{ls}(su, co, ls, ef)}{dls},$$

and the derivative of the SO_2 emission function with respect to limestone use is negative as specified in (6). Differentiating equations (A-1) and (A-4) a second time provides a similar proof for the second derivatives. Similar proofs also hold with respect to operating efficiency, sulfur content, and coal use.

Proof 2:

In order for plants to sell their gypsum by-product it is necessary that $NP_G \geq -C_G$. In the event that $NP_G = -C_G$, plants are indifferent between selling and not selling, and the quantity of sorbent used is equivalent in both states. Assuming $NP_G > -C_G$, the first order condition in the contract state evaluated at ls^{NC} is given formally as:

(A-7). $$\frac{\partial \pi^C}{\partial ls} = NP_G * \frac{\partial G(su, co, ls^{NC}, ef)}{\partial ls} - P_{ls} - P_{SO_2} * \frac{\partial SO_2(su, co, ls^{NC}, ef)}{\partial ls} > 0.$$

Equation (A-7) reveals that the first order condition (FOC) for profit maximization in the contract state is violated at the sorbent use level given by ls^{NC}. Using the curvature and monotonicity conditions for gypsum production and SO_2 emissions with respect to limestone use from equations (5) and (6) results in the following second order condition for profit maximization with respect to limestone use:

(A-8). $$\frac{\partial^2 \pi^C}{\partial ls^2} = NP_G * \frac{\partial^2 G(su, co, ls, ef)}{\partial ls^2} - P_{SO_2} * \frac{\partial^2 SO_2(su, co, ls, ef)}{\partial ls^2} \leq 0.$$

When $NP_G \geq 0$, the SOC given by equation (A-8) is satisfied for all levels of sorbent usage. Plants with a contract to sell gypsum choose a sorbent level, ls^C, such that $ls^C \geq ls^{NC}$, and equality of sorbent use between contract and no contract states holds only in the event that plants are indifferent between selling and not selling. On the other hand, the second derivative of the profit equation is not strictly negative when $NP_G < 0$, and second order conditions for profit maximization result in the following inequalities that must be satisfied at the maximum for the contract and no contract states, respectively:

(A-9). $$P_{SO_2} * \frac{\partial^2 SO_2(su, co, ls, ef)}{\partial ls^2} \geq NP_G * \frac{\partial^2 G(su, co, ls, ef)}{\partial ls^2},$$

(A-10). $$P_{SO_2} * \frac{\partial^2 SO_2(su, co, ls, ef)}{\partial ls^2} \geq -C_G * \frac{\partial^2 G(su, co, ls, ef)}{\partial ls^2}.$$

Recall that in order for a plant to sell gypsum it is necessary that $NP_G \geq -C_G$. As a result, if equation (A-10) is satisfied at ls^{NC}, then equation (A-9) is also satisfied at ls^{NC} for gypsum sellers, and profit maximizing sorbent use in the contract and no contract states still satisfies the inequality $ls^{C} \geq ls^{NC}$. Similar proofs hold with respect to scrubber operating efficiency, sulfur content, and coal consumption. The proofs on sulfur content and coal consumption hold if the coal price function is negatively sloped and convex, and the electricity generation function is positively sloped and concave, respectively.

Table III-A1: Tests for Common Trend in SO_2 Emissions[a]

	Estimated Coefficients (Std. Errors)			
Variable Name	**Net Effect (2yr lag)**	**Net Effect (4yr lag)**	**Net Effect (6yr lag)**	**Net Effect (8yr lag)**
Panel A: Dependent Variable is Log SO_2 Emissions				
Ln Total Heat	1.027***	1.026***	1.024***	1.025***
	(0.124)	(0.124)	(0.124)	(0.122)
Ln Total Ash	0.089	0.091	0.094	0.093
	(0.072)	(0.073)	(0.078)	(0.075)
Pseudo Contract	-0.064	-0.020	0.001	-0.004
	(0.072)	(0.076)	(0.094)	(0.097)
Number of obs.	953	953	953	953
R-squared	0.491	0.490	0.490	0.490
Number of boilers	94	94	94	94
Panel B: Dependent Variable is SO_2 Emissions				
Total Heat	501.982***	503.875***	505.141***	523.581***
	(67.962)	(68.257)	(71.668)	(74.736)
Total Heat Squared	$-2.64e^{-6}$***	$-2.67e^{-6}$***	$-2.69e^{-6}$***	$-2.89e^{-6}$***
	($6.52e^{-7}$)	($6.36e^{-7}$)	($6.53e^{-7}$)	($6.21e^{-7}$)
Total Ash	-0.011	-0.011	-0.011	-0.012
	(0.007)	(0.007)	(0.007)	(0.008)
Total Ash Squared	$1.45e^{-8}$***	$1.45e^{-8}$***	$1.46e^{-8}$***	$1.64e^{-8}$***
	($5.12e^{-9}$)	($5.08e^{-9}$)	($5.41e^{-9}$)	($5.95e^{-9}$)
Pseudo Contract	-2,244.201	-2,139.997	-2,565.465	-2,769.808
	(1,969.666)	(2,294.241)	(2,356.157)	(2,244.052)
Number of obs.	956	956	956	956
R-squared	0.318	0.323	0.329	0.331
Number of boilers	94	94	94	94

[a] Statistical significance at the 1%, 5%, and 10% level are represented by ***,**,and *, respectively. Although not reported, the models presented also include a full set of boiler and year fixed effects, scrubber type dummy variables, and scrubber manufacturer dummy variables.

Table III-A2: Alternative Specification Tests[a]

Model	Incremental Effect of Contract on SO_2 Emissions (Net Effect)	Incremental Effect of Contract on SO_2 Emissions (Efficiency)
Log-log	-1,488.412***	-1,446.466***
Linear	-1,715.224**	-1655.281**
Box-Cox	-1,448.411***	-1,411.396***
Tobit	-1,828.308***	-821.722*
Log-log-IV	-3,494.104***	-3,317.311**
Linear-IV	-4,254.094**	-3,921.203*

[a] Statistical significance at the 1%, 5%, and 10% level are represented by ***,**,and *, respectively. Incremental effects in non-linear log-log and Box-Cox models are estimated using a smearing estimate (Duan, 1983).

Table III-A3: First-Stage Estimates of Gypsum Sales Contracts[a]

Variable Name	IV Net Effect – Total SO_2 Emissions is dependent variable	IV Efficiency – Total SO_2 Emissions is dependent variable
Ln Total Heat	-0.17387***	-0.18726***
	(0.060)	(0.068)
Ln Total Ash	0.07261**	0.06155*
	(0.029)	(0.036)
Ln Total Sulfur	-----	0.01241
		(0.031)
Ln Total Sorbent	-----	-0.01179
		(0.010)
Ln Hours In Service	-----	0.03099
		(0.025)
Fifty Miles	0.39711***	0.39824***
	(0.081)	(0.081)
Hundred Miles	-0.61451***	-0.61884***
	(0.119)	(0.120)
Two Hundred Fifty Miles	0.00464	0.00849
	(0.057)	(0.055)
Hundred Miles Count	0.56572***	0.56752***
	(0.118)	(0.119)
Two Hundred Fifty Miles Count	0.04487*	0.04265*
	(0.023)	(0.023)
Number of obs.	1,530	1,527
R-squared	0.421	0.422
Number of plants	153	153
F-Statistic	18.780	18.623
F-Stat. p-value	(0.000)	(0.000)

[a] Statistical significance at the 1%, 5%, and 10% level are represented by ***,**,and *, respectively. Although not reported, the models presented also include a full set of boiler and year fixed effects, scrubber type dummies, and scrubber manufacturer dummies.