EXAMINING SPATIAL ARBITRAGE: EFFECT OF ELECTRONIC COMMERCE AND ARBITRAGEUR STRATEGIES

A Dissertation

Presented to

The Academic Faculty

By

Hemang C Subramanian

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in the

Scheller College of Business

Georgia Institute of Technology

August 2015

EXAMINING SPATIAL ARBITRAGE: EFFECT OF ELECTRONIC COMMERCE AND ARBITRAGEUR STRATEGIES

Approved By:

Dr Eric Overby, chair
Scheller College of Business
Georgia Institute of Technology

Dr. Saby Mitra,
Scheller College of Business
Georgia Institute of Technology

Dr. Sridhar Narasimhan
Scheller College of Business
Georgia Institute of Technology

Dr. Chris Forman
Scheller College of Business
Georgia Institute of Technology

Dr. Sam Ransbotham
Carroll School of Management
Boston College

Date: 6 July 2015

To

My parents, teachers, Saranya and Abyukta

ACKNOWLEDGEMENTS

This dissertation is the result of years of guidance and support from my professors, colleagues and family.

I am grateful to my dissertation committee chair and advisor Dr. Eric Overby for the exceptional inspiration, support, guidance and mentorship throughout my research program. I am grateful to Dr. Saby Mitra for his valuable guidance and encouragement at every step of my research over the past 5 years. I am grateful to Dr. Chris Forman, for his encouragement and support during the Ph.D. program the Scheller College of Business. I am grateful to Dr. Sridhar Narasimhan, for sharing his wisdom and collaboration on various technical projects during my research. I am grateful to Dr. Sam Ransbotham for being a great collaborator and a true inspiration. I am thankful to (late) Dr. Sandra Slaughter for her encouragement and mentorship over years and for setting such high standards of research scholarship. I whole heartedly thank Dr. Niculescu for guidance on various occasions during the Ph.D. program. Similarly, I thank Dr. D. J. Wu for the ITM seminar and for guidance through the years. I thank Dr. Han Zhang, Dr. Jeffrey Hu, Dr. Lizhen Xu and Dr. Mike Smith for their invaluable guidance on many occasions. I thank Dr. Jonathan Clarke whose advice and consultation on the literature in behavioral finance has helped me in my research thinking. In addition, I am thankful to numerous scholars in Information Systems for creating a wonderful body of knowledge which makes the pursuit of a career in research and teaching very exciting. I am indeed standing on the *"Shoulders of Giants"*. I thank my professors Dr. Frank Rothaermel, Dr. Anna M. Conti, Dr. Marie Thursby, Dr. Jerry Thursby, Dr. Ajay Kohli, Dr. Gautam Chalagalla, Dr. Debbie Turner, and Dr. Jay Lee for Ph.D. level courses that have shaped my research thinking.

The Information Technology Support group at the Scheller College of Business consisting of Mike Sewell, James Emery, Sheryl George, Stephen Lester and William Thomas has supported me immensely with computing needs for my research. I am grateful to my colleagues Mike Frutiger, Tina Xu, Karthik, Peng, Hongchang, Patricia, Zhanfei, Yang, Rohan and Narendra at the Scheller College of Business. I am grateful to my seniors Drs. German Retana, Tianshi Wu, Wen Wen, Peng, Chih-Hung and Denny for their support. I am grateful to Ms. Michelle Graham from administration for her prompt and generous help on many occasions.

Outside the college, my parents Dr. C. Subramanian and Mrs. Sulochana S. have been extremely supportive. Their wishes have been a guiding force for my life in every activity I pursue. My in-laws Mr. Krishnan V. and Mrs. Sajitha K have been very supportive of my career choice in research and teaching. I thank my brother Dr. Aneesh C S(and his wife Dr. Donata), sister Dr. Susmitha (and her husband Mr. Sudhakaran), my sister-in-law Silpa (and her husband. Sreejith C. R.) and brother-in-law Sarath for their support. The smiles and tantrums of my 3.5 year old daughter Abyukta have many a times given me the needed inspiration and energy for my research. The laughter of my little nephew Abhinav and nieces Anagha and Shreya have kept me inspired throughout my journey. I am grateful to my ex-colleagues Dr. Raghunath , Abhinav A, Dr. Arun, Dr. Nithin, Manjunath B.S and Neal Sato for their encouragement at various stages of my career and life so far. My research was made possible because of my wife Saranya's support and many sacrifices during this journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Markets increase social welfare by matching willing buyers and sellers. It is important to understand whether markets are fulfilling their societal purpose and are operating efficiently. The prevalence of spatial arbitrage in markets is an important indicator of market efficiency. The two essays in my dissertation study spatial arbitrage and the behaviors of arbitrageurs

Electronic commerce can improve market efficiency by helping buyers and sellers find and transact with each other across geographic distance. In the first essay, we study the effect of two distinct forms of electronic commerce on market efficiency, which we measure via the prevalence of spatial arbitrage. Spatial arbitrage is a more precise measure than price dispersion, which is typically used, because it accounts for the transaction costs of trading across distance and for unobserved product heterogeneity. Studying two forms of electronic commerce allows us to examine how the theoretical mechanisms of expanded reach and transaction immediacy affect market efficiency. We find that electronic commerce reduces the number of arbitrage opportunities but improves arbitrageur's ability to identify and exploit those that remain. Overall, our results provide a novel and nuanced understanding of how electronic commerce improves market efficiency. Studying arbitrageur strategies will help us understand how arbitrageur behaviors impact markets by increasing/reducing spatial arbitrage.

In the second essay, we study specialization strategies of arbitrageurs. Arbitrageurs specialize on asset type and sourcing locations. We investigate the role of specialization and find that specialization affects both arbitrage profits and arbitrage intensity. Subsequently, we find that specialization strategies evolve over time and different groups of arbitrageurs adapt differently based on behavioral biases and environmental factors. Overall, our findings support

the predictions of the adaptive markets hypothesis and help us understand antecedents such as capital, arbitrage intensity, etc. which affect the evolution of arbitrageur strategy.

CHAPTER 1: INTRODUCTION

Markets increase social welfare by matching buyers and sellers and enabling trade between buyers and sellers. However, impediments to trade caused by factors such as difficulties in transporting goods and/or government regulations inhibit free trade across geographic distances. Arbitrageurs exploit such opportunities in markets by purchasing goods at locations where the good is priced low; they then transport the good to another location to sell it at a higher price. This is known as spatial arbitrage. The law of one price and the condition for market efficiency is that prices of the same good across locations should be the same. (Transaction costs can vary, i.e., the cost of transporting the good between two locations and other fees associated with the good.) Spatial arbitrage and arbitrageurs are fundamental to market efficiency since arbitrageurs enable the matching of supply and demand across locations and, therefore, enable equilibration across geographic distances (Barrett 2008, Takayama and Judge 1964).

One of the most enduring streams of literature in Information Systems and Economics is the study of the effect of electronic commerce on markets and market efficiency. Several studies have shown how electronic commerce plays a major role in reducing price dispersion, thereby improving market efficiency (Bakos 1998, Brynjolfsson and Smith 2000). In most studies regarding market efficiency, scholars have used price dispersion as a measure of market efficiency. A key factor that has prevented scholars from studying spatial arbitrage in markets has been that of data limitations.

The first essay in this dissertation studies the effect of electronic commerce on market efficiency as indicated by spatial arbitrage. The study uses a unique dataset that identifies traders, locations and goods. We identify the effect of two different modes of electronic commerce (webcast and standalone electronic market) on spatial arbitrage and, by extension, market efficiency. We study differences in the effects of two electronic-trading channels based on channel features—namely, transaction immediacy and channel reach. Our results indicate that although electronic commerce reduces spatial arbitrage (and improves market efficiency), this effect is nuanced because of the variations in channel features between the two electronic commerce channels. Our results indicate that the webcast channel reduces the number of vehicles that are arbitraged, and the standalone electronic market channel increases the likelihood of exploiting arbitrage opportunities. Our findings indicate a nuanced effect, these two electronic channels – while reducing the overall opportunities in the markets also improve the ability of arbitrageurs to exploit market opportunities that remain.

In the third chapter of this dissertation, we study arbitrageur behavior with respect to the arbitrageur's specialization strategy. Recent advances in behavioral finance, which originated in the "Limits of Arbitrage," have attempted to show that arbitrage cannot entirely eliminate market inefficiencies because of limitations faced by arbitrageurs (Barberis and Thaler 2003, Mullainathan and Thaler 2000, Shleifer and Vishny 1997). Risk-avoidance, capital limitations, cognitive limitations, bounded rationality, and arbitrageur specialization are some factors identified in prior literature as factors limiting the arbitrageur.

A more recent stream of research called the Adaptive Markets Hypothesis provides a framework to explain predictions of behavioral finance (Lo 2004). This theoretical framework uses theories of evolution to explain behaviors of arbitrageurs (and other traders) in markets. The adaptive markets framework explains that behaviors of arbitrageurs evolve due to a combination of environmental factors and behavioral biases exhibited by arbitrageurs. One of the key predictions of this stream of research is that groups of traders will adapt their behaviors to environmental changes in specific ways and will evolve similarly. Based on the predictions of the limits of arbitrage and adaptive markets hypothesis, we empirically examine specialization strategies of arbitrageurs. Firstly, we test the theoretical predictions regarding arbitrageur specialization and the effects of specialization on arbitrage profits. Then, we analyze how different groups of arbitrageurs evolve behaviorally in markets with respect to their specialization strategies. Finally, we conclude by analyzing the antecedents that affect the evolution of arbitrageur strategies.

This dissertation is organized as follows. In Chapter 2, *Spatial Arbitrage, Electronic Commerce, and Market Efficiency,* we study the effect of electronic commerce on market efficiency as indicated by spatial arbitrage. In Chapter 3, *Spatial Arbitrage and Arbitrageur Specialization Strategies,* we examine the specialization strategies of arbitrageurs with respect to their choice of asset and sourcing locations. Finally, in Chapter 4 we conclude the dissertation by summarizing the key findings of Chapters 2 and 3.

Keywords: electronic commerce, spatial arbitrage, market efficiency, quasi-natural experiment, limits of arbitrage, discrete choice models, mixed logit, coarsened exact matching, behavioral bias, Gini coefficient, 3-stage least squares estimation

CHAPTER 2: SPATIAL ARBITRAGE, ELECTRONIC COMMERCE AND MARKET EFFICIENCY

2.1 Introduction

Markets can increase social welfare by matching willing buyers and sellers (McMillan 2002). However, geography can limit how efficiently markets match buyers and sellers, because it may be difficult for them to find and trade with each other across distance, even if such a trade would be optimal for both parties. Electronic commerce should facilitate trading across geographic distance in at least two ways: a) by improving the visibility of buyers and sellers in different geographic locations, and b) by eliminating the need for collocation between buyers and sellers (Bakos 1991). As such, studying the effect of electronic commerce on market efficiency is an important stream in information systems research and as well as in economics (e.g., Brynjolfsson and Smith 2000, Ghose and Yao 2011, Jensen 2007). We contribute to this stream by posing the following research question: How does electronic commerce affect market efficiency, as measured by the prevalence of spatial arbitrage?

Addressing this question allows us to make the following contributions. First, scholars typically use price dispersion to measure market efficiency. In contrast, we use the prevalence of spatial arbitrage, which is a more precise measure. Following Coleman (2009), we define spatial arbitrage as the purchase and subsequent resale of the same product in different geographic locations to exploit a price discrepancy.1 If buyers and sellers do not match efficiently across locations, then this creates spatial arbitrage opportunities in which a third party – an arbitrageur – purchases products from sellers in

1 To limit definitional confusion, we do not consider instances in which buyers (sellers) eschew buying (selling) at one location in favor of another because of price differences to be "arbitrage", even though some authors use the term that way (e.g., Jensen 2007). This is because arbitrage, as we define it and as is consistent with the textbook definition (Sharpe et al. 1995, p. 1001), requires both a purchase and a sale.

locations where prices are low and resells them to buyers in locations where prices are high. Spatial arbitrage will occur as long as the transaction costs associated with moving products between locations are lower than the price difference between locations, becoming more prevalent as the gap between the transaction costs and the price difference widens. This measure has several advantages over price dispersion, including inherently accounting for transaction costs and for unobserved product heterogeneity. To illustrate the first advantage, consider a perfectly efficient market, i.e., one in which buyers and sellers match optimally, regardless of geographic location. According to the law of one price, this market might have substantial price dispersion, because prices in an efficient market can vary up to the transaction costs of moving products between locations (Persson 2008). But there would be no spatial arbitrage in this market; indeed, "no arbitrage" is a classic condition for market efficiency (Barrett 2008; Takayama and Judge 1971). Thus, a researcher using price dispersion to measure market efficiency might incorrectly conclude that this market was inefficient. A researcher using the prevalence of spatial arbitrage to measure market efficiency (and not finding any) would not make this mistake. Our use of this new measure is important because improving measurement is fundamental to scientific advancement.

Second, we examine *why* electronic commerce affects market efficiency by examining two theoretical factors that distinguish electronic commerce from traditional commerce: reach and transaction immediacy. Reach allows traders to find and transact with each other across geographic distance, and transaction immediacy allows them to conduct transactions immediately, at any time. Empirically, we study two distinct forms of electronic commerce, both of which provide expanded reach but only one of which provides transaction immediacy. This distinction allows us to better understand the mechanisms behind the effect of electronic commerce. It also allows us to contribute to the growing body of research that recognizes that not all forms of electronic commerce are the same (Ghose et al. 2013).

Third, we contribute to the empirical literature on arbitrage. Arbitrage is a central mechanism in many foundational economic theories such as the law of one price and the efficient markets hypothesis. In these theories, arbitrageurs are the critical agents who identify and exploit market inefficiencies as they arise, thereby restoring efficiency by rebalancing supply and demand (Shleifer and Vishny 1997). Despite arbitrage's central place within theory, there is little empirical evidence about how arbitrageurs behave. We contribute to this literature by studying how arbitrageurs choose where to source products for arbitrage, including how this is affected by electronic commerce.

Despite its advantages, the prevalence of spatial arbitrage has rarely been used to measure market efficiency. This is because spatial arbitrage transactions are difficult to observe: a researcher must be able to observe a trader i who purchases an item at "source" location k and then quickly resells the *same* item at a different "destination" location l. This requires unique (and consistent) trader, location, and item-level identifiers. We overcome this by studying spatial arbitrage in the context of the wholesale used vehicle market, where we track the trading history of each vehicle based on its unique Vehicle Identification Number (VIN). Our data also contain unique and consistent identifiers for the traders and market locations. Another advantage of this context is that this market has implemented two distinct electronic channels: a webcast channel that allows electronic access to the traditional physical market and a standalone electronic market. Both channels provide expanded reach, but only the standalone electronic market provides transaction immediacy. These channels accounted for an increasing number of transactions over our sample period (2003 to 2010).

Theoretically, the expanded reach provided by both channels should help "regular" buyers purchase directly from sellers in remote locations, thereby disintermediating the arbitrageurs and reducing arbitrage opportunities. But it should also help arbitrageurs find and exploit previously hidden opportunities. The transaction immediacy provided by the standalone electronic market should also help arbitrageurs identify and exploit arbitrage opportunities before they dissipate. This suggests a nuanced effect in which electronic commerce eliminates many arbitrage opportunities but improves arbitrageurs' ability to identify and exploit those that remain. Whether each of the channels increases or decreases the number of arbitrage transactions (i.e., exploited opportunities) depends on whether the channel increases the efficiency with which arbitrageurs exploit opportunities more than it reduces the number of opportunities. This is not clear a priori and warrants empirical testing.

Empirically, we find support for the nuanced effect: the number of arbitrage opportunities decreased as electronic commerce became more widely used, but the percentage of arbitrage opportunities that were exploited increased. We examined this further by leveraging the phased implementation of the webcast channel across the market to conduct a quasi-natural experiment to examine its effect. We used a similar matching estimation to examine the effect of the standalone electronic market. The webcast channel has a negative effect on spatial arbitrage, whereas the standalone electronic market has a positive effect. As theorized, both channels increased geographic purchasing reach. Also, buyers leveraged the transaction immediacy of the standalone electronic market to exploit arbitrage opportunities by quickly identifying and purchasing undervalued vehicles. (They did not do this via the webcast channel, because it does not provide transaction immediacy.) We conclude that the "opportunity exploitation" effect outpaced the "opportunity reduction" effect in the standalone electronic market (yielding the positive overall effect) but not in the webcast channel (yielding the negative overall effect). Because the webcast channel was more widely used during the sample period, the "opportunity reduction" effect outpaced the "opportunity exploitation" effect overall, resulting in an overall decline in the number of arbitrage transactions (i.e., exploited opportunities). Consistent with the negative effect of the webcast channel on arbitrage, we find that arbitrageurs' preference of facility from which to source vehicles declined with the degree to which the webcast channel had been deployed. In general, we find that arbitrageurs prefer to source vehicles at locations that are difficult for other traders to access (both physically and electronically), likely because these locations are more isolated from market-wide price trends.

Overall, our results provide a fuller and more nuanced picture of the way in which electronic commerce affects market efficiency than has previously been documented (to our knowledge). Our analysis suggests that electronic commerce can improve efficiency in two ways. The first way is that electronic commerce can help buyers and sellers trade with each other across geographic locations, thereby providing better supply/demand balance. This is consistent with prior research (e.g., Aker 2010, Jensen 2007). The second way – which is more subtle and has not been empirically shown to our knowledge – is that electronic commerce can help arbitrageurs better exploit any remaining supply/demand imbalances, the very act of which helps return a market to efficiency when it strays.

In §2, we review the prior literature and describe the differentiators of our study. In §3, we present our theory and hypotheses. In §4, we describe our empirical setting. In §5, we present our analyses and results. In §6, we conclude with a summary of the paper's findings, contributions, and limitations.

2.2 Literature review

Our research contributes to two research streams: a) the literature on market efficiency, including how it is affected by electronic commerce, and b) the literature on spatial arbitrage and arbitrageur behavior.

2.2.1 Research on market efficiency and electronic commerce

Scholars have typically measured market efficiency by examining price dispersion (e.g., Chellappa et al. 2011, Ghose and Yao 2011). The intuition is that a high degree of price dispersion indicates that products are not being allocated efficiently, i.e., that supply is too high for the demand in some regions and too low in other regions. This will lead to low prices in the former regions and to high prices in the latter regions, creating price dispersion. Scholars have also measured market efficiency by analyzing the co-movement of prices at different locations over time, which is essentially an analysis of how price dispersion evolves longitudinally (Alexander and Wyeth 1994). The intuition for this measure is that prices in an efficient market can differ across locations up to the cost of transport between locations (Persson 2008, Takayama and Judge 1971), but this difference should be relatively constant over time. I.e., prices across locations should move up or down together, such that price dispersion remains mostly unchanged. If prices do not move together, then this might reflect excess supply or demand in certain locations at certain times, reflecting the market's inefficiency at balancing supply and demand.

To assess the effect of electronic commerce on market efficiency, scholars have examined whether price dispersion is lower online than offline (Brown and Goolsbee 2002, Brynjolfsson and Smith 2000). A common motivation for these studies is that because transaction costs are lower online than offline (as is typically assumed), supply and demand will be more efficiently distributed online, resulting in lower price dispersion online. Empirical support for this is mixed (Baye et al. 2006). Another approach scholars have used is to examine whether price dispersion in a market declines as electronic commerce is adopted. Studies using this approach have shown that electronic commerce has led to reduced price dispersion across locations (e.g., Aker 2010, Overby and Forman 2015, Parker et al. 2013).

We extend this literature in two ways. First, we use the prevalence of spatial arbitrage, rather than price dispersion, to measure efficiency. Second, we examine two distinct forms of electronic commerce.

2.2.1.1 Measuring efficiency via the prevalence of spatial arbitrage rather than via price dispersion:

Despite its widespread use, price dispersion has two key limitations as a measure of efficiency (Badiane and Shively 1998, Barrett 2008). First, a "baseline" level of price dispersion may exist in a perfectly efficient market, with this baseline equal to the transaction costs of moving products between locations (Takayama and Judge 1971). Without knowledge of these costs, it is difficult to determine what level of price dispersion represents the baseline level vs. what level represents inefficiency in matching buyers and sellers. Prior research that has relied on price dispersion to measure efficiency has often worked around this issue by studying price dispersion on a relative basis. E.g., if the level of price dispersion in a market declines over time, then it is reasonable to assume that efficiency has improved. However, this approach makes the (potentially erroneous) assumption that the baseline does not also decline.[2] Second, price dispersion will be an inaccurate measure of market efficiency if products whose prices are being compared are not comparable due to unobserved factors such as differences in quality. Although this is not a concern for perfectly homogeneous products, it is a concern for any product whose attributes may vary from unit-to-unit (e.g., agricultural crops, automobiles, crude oil, metals, etc.)

To illustrate the first limitation, consider (hypothetically) that the price of a barrel of crude oil is $45 in the United States ("US") and $55 in the United Kingdom ("UK"). On one hand, this $10 price dispersion might reflect an inefficient imbalance in supply and demand (e.g., too many buyers in the UK). On the other hand, supply and demand might be perfectly balanced, with the $10 dispersion resulting from different tax/tariff regimes and/or non-zero transport costs between the US and the UK. In the first case, buyers in the UK (sellers in the US) would be better off shifting their demand (supply) to the US (UK) where they could exploit the inefficiency for profit. In the second case, they would not, because there is no inefficiency to exploit. To illustrate the second limitation, consider that crude oil in the UK might be of higher quality than that in the US. In this case, the $10 price dispersion would reflect this quality difference, not inefficiency in balancing supply and demand.

Using the prevalence of spatial arbitrage as the measure of market efficiency addresses these two limitations. First, the arbitrage measure inherently accounts for the transaction costs of moving products between locations. To elaborate, consider that although it is difficult for researchers to assess whether a given level of price dispersion exceeds the transaction costs of moving products between locations, it is far less difficult for arbitrageurs. This is because arbitrageurs are market specialists who are keenly aware

2 For example, assume that price dispersion between two locations over a four-year period is 20, 19, 18, and 17. This decline could be due to improved efficiency. But it could also be due to an annual 1-unit reduction in the costs of transporting products between locations – with efficiency remaining unchanged. Or, if transport costs reduced by 2-unit annually, then efficiency could actually be decreasing.

of price disparities and the transaction costs associated with exploiting them (Shleifer and Vishny 1997). Arbitrageurs consider the transaction costs when determining whether to engage in arbitrage. If the transaction costs are lower than the price disparities, then arbitrage will occur, becoming more prevalent as the gap between the transaction costs and the price disparities widens, i.e., as the market becomes more inefficient. If the transaction costs are higher than the price disparities, then arbitrage will not occur. Essentially, using the prevalence of spatial arbitrage to measure efficiency eliminates the burden from the researcher of trying to estimate whether a given level of price dispersion represents inefficiency. Instead, this question is answered (directly) by the arbitrageurs. Second, the arbitrage measure is immune to potential quality differences across products, because the same product is traded in both locations (unless the product is altered during transport).

To illustrate these advantages, we return to the crude oil example. If we observe arbitrage activity in which arbitrageurs purchase in the US for $45 per barrel and resell in the UK for $55, then we can safely assume that the transaction costs are less than $10 per barrel and that the market is inefficient. Also, because the same oil is being transacted in both locations, we can eliminate the possibility that the $10 price dispersion is due to quality differences. On the other hand, if we do not observe arbitrage activity, then the $10 price dispersion is more likely to reflect non-zero transaction costs or differences in quality than inefficiency. Table 1.1 summarizes the advantages of measuring market efficiency via the prevalence of spatial arbitrage instead of price dispersion.

<u>2.2.1.2 Examining two distinct forms of electronic commerce</u>*:* A limitation of research on electronic commerce and market efficiency is that it tends to implicitly assume that all forms of electronic commerce are the same. Prior studies have either examined a single form of electronic commerce and used that to draw conclusions about electronic commerce in general (e.g., Jensen, 2007) or studied electronic commerce in general terms without differentiating between different types of electronic commerce (e.g., Brynjolfsson and Smith, 2000). A differentiator of our study is that we examine the

effects of two distinct forms of electronic commerce: a webcast channel that permits electronic access to the traditional physical market and a standalone electronic market. This helps us examine the mechanisms responsible for the effect of electronic commerce. To elaborate, different forms of electronic commerce have different features. Assume that form #1 of electronic commerce has feature A and no effect on market efficiency, while form #2 has features A and B and a positive effect on market efficiency. This would suggest that feature B (and not feature A) is the reason for the effect. Such a conclusion would not be possible if only one of these two forms of electronic commerce were examined. Our analysis of multiple forms of electronic commerce also contributes to the growing literature that recognizes that different types of electronic commerce (e.g., PC vs. phone-based) may have different effects (e.g., Ghose et al. 2013). Table .12 summarizes how our research is distinct from prior research on the effect of electronic commerce on market efficiency.

Table 2.1: Issues with measuring market efficiency via price dispersion and how those are remedied by using the prevalence of spatial arbitrage as the measure.

	Measure of Market Efficiency	
	Price Dispersion	Prevalence of Spatial Arbitrage
Issue #1: Transaction costs of trading across locations	Does not account for these costs, making it difficult to tell if a given level of price dispersion represents inefficiency or not.	Accounts for these costs automatically because arbitrageurs consider them when determining whether to engage in arbitrage.
Issue #2: Product quality differences between locations	Only accounted for if the researcher compares homogeneous products or adjusts for product quality differences.	Accounts for this automatically because the same product is traded at both locations.

Table 2.2: Categorization of research on the effect of electronic commerce on market efficiency, including differentiators of the present study.

		Measure of Market Efficiency	
		Price Dispersion	Prevalence of Spatial Arbitrage
Analysis of different forms of electronic commerce?	Yes		*Present study*
	No	Aker (2010), Brown and Goolsbee (2002), Chellappa et al. (2011), Jensen (2007), etc.	Overby and Clarke (2012)

2.2.2 Research on spatial arbitrage and arbitrageur behavior

Another contribution of our study is that we document new empirical findings about spatial arbitrage and arbitrageur behavior. This is important because despite arbitrage's central role in theory such as the law of one price and the efficient markets hypothesis, it is difficult to observe. As a result, much of what we believe about arbitrage is based on maintained assumptions that have not been subjected to empirical testing. Empirical analysis of spatial arbitrage is rare for two main reasons. First, because spatial arbitrage happens relatively infrequently and constitutes a small fraction of a market's trade volume, empirical analysis of spatial arbitrage requires large datasets. Second, observation of spatial arbitrage requires unique (and consistent) identifiers for individual products, traders, and locations so that the trading history of products can be tracked. Because these data are often unavailable, scholars have used price dispersion (as noted above) to infer the presence of spatial arbitrage opportunities. A drawback to this approach is that it tells us very little about how arbitrageurs behave. We help address this situation and contribute to the research on arbitrage by studying how arbitrageurs choose where to source products, including how electronic commerce affects this.

There is one other study of which we are aware that examines spatial arbitrage at a transaction-level. Overby and Clarke (2012) used data from the wholesale used vehicle industry to analyze how sellers' bounded rationality causes them to distribute some vehicles sub-optimally, thereby creating opportunities to spatially arbitrage those vehicles. We depart from their analysis in several ways. First, we contribute novel findings about how arbitrageurs behave by examining where they choose to source products, which – despite the centrality of arbitrage to foundational economic theory – has not been previously examined to our knowledge. Second, we analyze the longitudinal change in both the number of arbitrage opportunities and the number of arbitrage transactions (i.e., the number of exploited opportunities); Overby and Clarke analyzed only the latter. This allows us to examine how electronic commerce affects how efficiently arbitrageurs exploit arbitrage opportunities. Third, Overby and Clarke reported a negative correlation between webcast channel use and spatial arbitrage. We corroborate this negative relationship by conducting a quasi-natural experiment, which is a stronger identification strategy than that used by Overby and Clarke, thereby improving our ability to attribute causality to the relationship. Fourth, we analyze how both the webcast channel and the standalone electronic market affect spatial arbitrage. The standalone electronic market was ignored by Overby and Clarke, who – as in the studies mentioned above – studied only a single form of electronic trading. We find that the standalone electronic market has a *positive* effect on spatial arbitrage. Fifth, our analysis of multiple types of electronic commerce allows us to explore the mechanisms by which electronic commerce affects spatial arbitrage. This allows us to examine not only whether electronic commerce affects spatial arbitrage activity but also why.

2.3 Theory and hypotheses

2.3.1 Effect of electronic commerce on spatial arbitrage

To consider the theoretical effect of electronic commerce on spatial arbitrage, it is useful to consider the features that distinguish electronic commerce from traditional commerce. We focus on two features for our analysis: a) reach, which is a spatial feature, and b) transaction immediacy, which is a temporal feature. First, we define reach as the ability for traders to find and transact with each other across geographic distance. Electronic commerce expands reach by lowering traders' search costs, such that they can more easily find and consummate trading opportunities with partners in remote locations. Second, electronic commerce typically provides transaction immediacy, which we define as the ability to conduct a transaction immediately, at any time (e.g., when physical stores are closed, without having to wait in line, etc.)[3] (Here, we are referring to the immediacy of the transaction itself, not necessarily to the

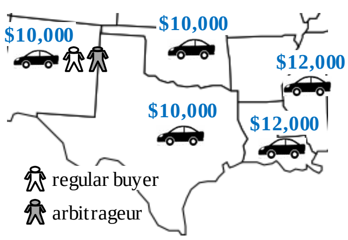
[3] We include the word "typically" because it is possible for an electronic commerce system to only allow transactions to be conducted at certain times. One of the electronic channels in our analysis operates this way.

immediacy of providing/receiving the product exchanged in the transaction.)

We next consider how these features of electronic commerce might affect spatial arbitrage. Spatial arbitrageurs are middlemen who purchase from a seller in location *k* and then resell to a buyer in location *l*. The reach provided by electronic commerce should help buyers and sellers in different locations find and transact with each other directly, thereby disintermediating the arbitrageur "middleman". This is consistent with recent research that shows that buyers use electronic commerce to shift demand from nearby sellers whose prices are high to more remotely located sellers whose prices are lower (Overby and Forman 2015). On the other hand, arbitrageurs can benefit from the reach of electronic commerce just as "regular" buyers and sellers can. For example, the reach provided by electronic commerce might help arbitrageurs identify locations across which supply and demand are imbalanced, thereby helping them identify and exploit previously hidden arbitrage opportunities. This suggests that electronic commerce will have a nuanced effect; it should reduce the number of arbitrage opportunities while simultaneously improving arbitrageurs' ability to identify and exploit the opportunities that remain. In other words, increasing use of electronic commerce should lead to a decrease in the number of arbitrage opportunities but to an increase in the percentage of arbitrage opportunities that are exploited.

The transaction immediacy of electronic commerce should also affect spatial arbitrage. Arbitrageurs are continuously looking for opportunities in which products in location *k* are priced below their value in location *l*. These opportunities are fluid and do not persist indefinitely, because they depend on dynamic supply/demand conditions in both locations. Thus, if an arbitrageur identifies a potential arbitrage opportunity between locations *k* and *l*, it is important for him to purchase the product from location *k* before conditions change. The reach provided by electronic commerce will help arbitrageurs find these opportunities, and the transaction immediacy of electronic commerce will allow arbitrageurs to purchase the products before their prices change or they are purchased by a "regular" buyer. Thus, the transaction immediacy of electronic commerce should enhance arbitrageurs' ability to exploit arbitrage opportunities.

“Regular” buyers could also leverage the transaction immediacy of electronic commerce to purchase undervalued products, which would take opportunities away from the arbitrageurs. However, we expect arbitrageurs to take fuller advantage of transaction immediacy to purchase undervalued products. This is because there are many instances in which a product will appear undervalued to an arbitrageur but not to a regular buyer. To see this, consider that a regular buyer is only interested in whether a product in a remote location is undervalued relative to his location, i.e., he goes online to buy more cheaply than he can buy locally. By contrast, arbitrageurs are interested in whether a product in *any* location is undervalued relative to *any other* location. Figure 2.1 provides an illustration. Also, research in behavioral economics suggests that arbitrageurs will identify and seize arbitrage opportunities more quickly than will regular buyers, because the former have the expertise and time to monitor market conditions more closely (Peng and Xiong 2006, Shleifer and Vishny 1997). The logic above suggests the following hypotheses.



The figure illustrates (hypothetically) a regular buyer (in white) and an arbitrageur (in gray), both located in New Mexico. The (hypothetical) average price for a 2012 Honda Accord is shown for different locations. The $10,000 vehicles in Oklahoma and Texas will not appear undervalued to the regular buyer, who will compare them to vehicles in New Mexico. By contrast, they will appear undervalued to the arbitrageur, who will compare them to vehicles not only in New Mexico but also in Louisiana and Mississippi.

Figure 2.1: Illustration of why more products will appear undervalued to arbitrageurs than to regular buyers.

H1a, H1b: Increasing use of electronic commerce is: a) negatively associated with the number of arbitrage opportunities, but b) positively associated with the percentage of arbitrage opportunities that are exploited.

Whether electronic commerce leads to an increase or decrease in the *number* of arbitrage opportunities that are exploited (i.e., the number of arbitrage transactions, aka the prevalence of spatial arbitrage) depends on whether the increased efficiency with which arbitrageurs exploit opportunities (i.e., the "opportunity exploitation" effect) outpaces the reduction in opportunities (i.e., the "opportunity reduction" effect). This should depend on the degree to which the form of electronic commerce supports expanded reach and transaction immediacy. In particular, assuming the degree of expanded reach is held constant, a form of electronic commerce that provides transaction immediacy is more likely to foster arbitrage activity (or at least inhibit it less) than a form that does not.

H2: Different forms of electronic commerce will have different effects on the prevalence of spatial arbitrage, with forms that support transaction immediacy having a more positive effect.

2.3.2 Factors that affect arbitrageur behavior of where to source products for later arbitrage

Several factors may influence the locations from which arbitrageurs choose to source products. First, arbitrageurs should prefer source locations where they expect to find a high percentage of inexpensive products that they can profitably arbitrage. Arbitrageurs should also prefer source locations at which prices for the products being offered vary widely across locations. This is because the more product prices vary across locations, the larger the potential arbitrage profits.

H3a, H3b: Arbitrageurs will prefer source locations at which: a) a high percentage of products sell for below their market value, and b) prices for the products being sold vary widely across locations.

Arbitrageurs should seek to limit the cost of sourcing vehicles (thereby increasing their arbitrage profit) by purchasing at locations close to them. However, if two locations are equidistant, an arbitrageur should prefer the location that is more difficult for other buyers/sellers to access, i.e., that is more isolated from the rest of the market. To explain, consider a location that is easily accessible to buyers/sellers. If prices at such a location become artificially low (high), then buyers (sellers) can easily shift demand (supply) there, thereby rebalancing supply and demand and eliminating would-be arbitrage opportunities. However, if this location was difficult for buyers/sellers to access, then prices would take longer to equilibrate to the market average, making the location an attractive source location.

H4a, H4b: Arbitrageurs will prefer source locations that are: a) nearby, but b) relatively difficult for other traders to access.

2.4 Empirical context

The empirical context for our study is the U.S. wholesale used vehicle market. Buyers in this market are used vehicle dealers. Most use the market as a source of inventory for their retail lots, while a small minority use the market to source vehicles for arbitrage within the wholesale market. The former (whom we will refer to as "buyers" or "regular buyers") try to make money from the difference between vehicles' retail and wholesale prices. The latter (whom we will refer to as "arbitrageurs") try to make money by exploiting inefficiencies within the wholesale market. Sellers are of two types: a) used vehicle dealers, and b) institutional sellers such as rental car firms. The former use the market either to sell vehicles that they do not sell retail or (much less commonly) to sell vehicles that they are arbitraging within the wholesale market. Institutional sellers use the market to sell vehicles retired from their fleets. Data were provided by an intermediary in the market that facilitates trades between buyers and sellers. The intermediary operates over 70 physical market facilities across the U.S. as well as the webcast channel and standalone electronic market described below. The data consist of 40,657,724 successful transactions facilitated by the intermediary from 2003 to 2010. Variables are described in Table 2.3.

Traditionally, the U.S. wholesale used vehicle market has functioned as a physical market in which buyers, sellers, and vehicles are collocated at market facilities. Each facility has a large parking lot for vehicle storage and a warehouse-type building equipped with multiple *lanes*, which are essentially one-way streets. Transactions occur as follows. Vehicles are driven down a lane – one at a time – where buyers interested in that vehicle will have gathered. An auctioneer solicits bids for each vehicle and awards the vehicle to the highest bidder, assuming he meets the seller's reserve price. This process takes

30-45 seconds, after which another vehicle is auctioned. It is common for vehicles to be auctioned in multiple lanes at the same facility concurrently.

2.4.1 Two distinct forms of electronic commerce

In the early 2000's, the intermediary who operates these facilities began simulcasting via the Internet the physical auctions as they were occurring at the facilities. This allows buyers to experience the live audio and video of the auctions via an Internet browser, and it permits them to bid on vehicles in competition with the buyers who are physically collocated at the facility. As such, this "webcast" channel provides buyers with electronic access to the auctions occurring in the physical market. The webcast channel was implemented in phases. This is because implementing the channel required that camera, microphone, and other equipment be implemented *in each lane at each facility*. This means that we observe many instances in which highly similar vehicles were auctioned at the same facility on the same day, some in lanes that were equipped for webcast and some in lanes that were not. As discussed below, we leverage this in a quasi-natural experiment to assess the effect of the webcast channel on the probability that a purchased vehicle is later arbitraged. Figure 2.2 shows how the percentage of vehicles available via webcast increased as the channel was deployed.[4] The intermediary also operates a standalone electronic market whose format is similar to that of eBay. In this market, sellers post listings of their vehicles, and buyers have the option to purchase them for a fixed "Buy Now" price or to bid for

4 We estimated the webcast implementation date for each lane/facility combination (i.e., lane 1 in Las Vegas, lane 2 in Las Vegas, etc.) as follows. The data denote whether a vehicle sold in the physical market was purchased by a buyer using the traditional physical channel or the webcast channel. For each lane/facility combination, we recorded the date of the first webcast purchase and used that as the webcast implementation date for that lane/facility. We considered all vehicles offered in that lane/facility from that date forward to be available via webcast. We used this to determine whether any given vehicle was available via webcast. Because there could be a lag between webcast implementation and the date of the first webcast purchase for a lane, we reran our analysis after adjusting the webcast implementation date by subtracting 1 week, 3 weeks, and 6 weeks. This does not affect our results.

them. Typically, a winning bidder exceeds the seller's (hidden) reserve price. But sometimes the seller will sell to the high bidder before the reserve price has been met. It is important to note that sellers choose to offer a vehicle either: a) at a physical market facility, where the vehicle is available to buyers present at that facility and to buyers accessing that facility via webcast (if enabled in the vehicle's lane), or b) in the standalone electronic market. Buyers can purchase a vehicle from: a) a physical market facility via the traditional physical channel, b) a physical market facility via the webcast channel (if enabled), or c) the standalone electronic market.

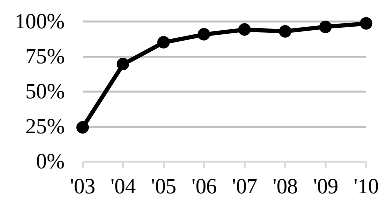


Figure 2.2: Annual percentage of vehicles offered in the physical market that were offered in webcast enabled lanes

From a theoretical perspective, the webcast channel and the standalone electronic market differ in their ability to provide expanded reach and transaction immediacy. Both forms of electronic commerce provide expanded reach: using either form, buyers can easily search for and purchase vehicles from remote locations without having to travel.5 However, due to their different designs, only the standalone electronic market provides transaction immediacy. This is because the webcast channel is an electronic access channel to auctions occurring in the physical market. As a result, the webcast channel can only be used to purchase vehicles during the 30-45 second window within which they are being auctioned at the

5 The channels provide expanded reach because they lower search costs. The webcast channel reduces search costs by letting buyers "look in" on and participate in auctions occurring across the country. This lowers buyers' cost of searching for and acquiring vehicle and price information across facilities. The webcast channel reduces buyers' search costs by aggregating vehicle listings from across the country in a single place.

physical market facility. Also, a buyer can only purchase a vehicle in the webcast channel if he places the highest bid and has this bid accepted by the seller. This means that the webcast channel cannot be used by a buyer to purchase a vehicle immediately, at any time; i.e., there is no transaction immediacy. By contrast, most vehicle listings in the standalone electronic market include a posted "Buy Now" price. Buyers can click this button at any time to purchase the vehicle immediately. We leverage these differences in our analysis, including to test H2.

Table 2.3: Description of variables.

Variable	Description	Descriptive statistics
Arbitraged	Denotes whether a purchased vehicle was spatially arbitraged (=1) or not (=0).	Mean: 0.0079 SD: 0.0885
FacilityID	The ID for the facility at which the vehicle was located.	-
FacilityZip	The facility's zip code.	-
LaneID	The ID for the lane at the facility in which the vehicle was auctioned.	-
TraderID	The ID for the trader. The same ID is used regardless of whether the trader is the buyer or seller in a transaction.	-
BuyerZip	The buyer's zip code. Could also be thought of as *ArbitrageurZip* when the buyer later arbitrages the vehicle.	-
BuyerDistance	The distance in miles between *BuyerZip* and *FacilityZip*.	Mean: 205; SD: 343
SellerType	Denotes the type of seller (institutional or dealer). 1=institutional; 0=dealer.	Mean: 0.57; SD: 0.50
SellerPct Arbitraged	The percentage of vehicles sold by the seller over the previous 28 days that were arbitraged. If over previous 28 days the seller sold 0 vehicles, we increased the number of days until the number sold was greater than 0.	Mean: 0.006 SD: 0.023
SaleDate	The date the vehicle was sold.	-
VIN	The vehicle's unique Vehicle Identification Number.	-
VehicleYear	The model year of the vehicle.	Mean: 2006.6; SD: 2.43
Make	The make of the vehicle, e.g., Chevrolet, Toyota, etc.	-
Model	The model of the vehicle, e.g., Tahoe, Camry, etc.	-
Price	The vehicle's sales price.	Mean: 10470; SD: 8060

Valuation	The vehicle's market value as estimated by the intermediary that provided the data.	Mean: 10589; SD: 8051
Mileage	The odometer reading of the vehicle.	Mean: 64302; SD: 49473
Standalone Electronic Market	Denotes whether a vehicle was offered in the standalone electronic market (=1) or the physical market (=0).	Mean: 0.02; SD: 0.14
Webcast Enabled	Denotes whether a vehicle in the physical market was offered in a webcast enabled lane (=1) or not (=0).	Mean: 0.76; SD: 0.44
WebcastBuyer	Denotes whether a vehicle offered in the physical market was purchased by a buyer using the webcast channel (=1) or the physical channel (=0).	Mean: 0.09; SD: 0.28
BuyFee	The transaction fee paid by the buyer.	Mean: 182.1; SD: 144.9
SellFee	The transaction fee paid by the seller.	Mean: 134.5; SD: 184.6
TransportFee	The transport fee between the facility and the buyer's location. Not available for all transactions; see the appendix.	Mean: 160.8; SD: 161.8
ListingDate	The date a vehicle was listed on the standalone electronic market.[a]	-
Transaction Type	Denotes whether a vehicle was purchased in the standalone electronic market via the Buy Now mechanism or auction (either before or after the hidden reserve price is met).[a]	Buy Now: 58%; Auction (after reserve met): 35%; Auction (before reserve met): 7%
[a] Not available for all transactions; see §5.2.3.		

2.4.2 Identification of spatial arbitrage

Following Overby and Clarke (2012), we identified instances of spatial arbitrage as follows. First, we identified what we refer to as "flips". A "flip" is a pair of transactions for the same vehicle (identified by its unique *VIN*) in which the buyer in the first transaction is the seller in the second transaction (as identified by his unique *TraderID)*. Flips occur when an arbitrageur is engaging in spatial arbitrage, but they may also occur for other reasons. For example, a dealer may flip a vehicle if he is unable to retail it and chooses to liquidate it in the wholesale market. Or, a dealer may flip a vehicle after making improvements to it (e.g. repairing dents, replacing tires). There are 2,749,524 flips in the sample. We delineate spatial arbitrage from other types of flips in two ways. First, we limit our focus to cross-facility flips, i.e., those in which the two transactions that comprise the flip occur at different facilities (as identified by the unique *FacilityIDs*). This is useful for delineation because a spatial arbitrage transaction must occur across different facilities (by definition), whereas flips attributable to other reasons are likely to be same-facility flips. For example, a dealer who is flipping a vehicle he changed/improved or failed to retail is likely to sell at the same facility from which he purchased in order to avoid the cost of transporting the vehicle to a different facility. Second, we only consider those flips completed within α days to be spatial arbitrage. We set α=7 in our primary analysis and varied α for robustness. The 7-day interval is reasonable because of the time needed to: a) complete paperwork at the source facility where the vehicle was purchased, b) transport the vehicle to the destination facility, and c) register the vehicle for sale at the destination facility. Increasing the α threshold increases the probability that we will falsely classify a flip as arbitrage, because a longer time period increases the probability that the vehicle has been changed/improved or that the dealer is liquidating a vehicle that he failed to retail. An example of an instance of spatial arbitrage is as follows: *TraderID* #111 purchased a vehicle with *VIN* 1B3EL36R54N976952 at the Miami facility on February 10, 2003 for $10,000 and then sold the vehicle at the New Orleans facility on February 13, 2003 for $11,500.

2.5 Analysis

2.5.1 Testing H1: The association between electronic commerce, the number of arbitrage opportunities, and the percentage of opportunities that are exploited

We estimated the number of spatial arbitrage opportunities in the market by assessing whether each vehicle j purchased at facility k on day t could have been profitably arbitraged at a different facility l within α days (we set α=7 and varied it for robustness). We did this as follows. First, we matched vehicles sold at each facility k (the source facility) on day t to vehicles sold at every other facility l (the destination facilities) between day $t+1$ and day $t+\alpha$. We matched on *VehicleYear*, *Make*, *Model*, *Mileage* (coarsened into bins of 1,000 miles), and *Valuation* (coarsened into bins of $1,000). Second, we estimated whether the vehicle from the source facility could have been profitably arbitraged at each matching destination facility by taking the mean *Price* of the matched vehicles at the destination facility and subtracting out *Price* at the source facility, the estimated transport cost between the facilities, and transaction fees (including the *BuyFee* at the source facility and the mean *SellFee* at the destination facility).[6] If this difference was more than $500 (after adjusting for inflation), we counted this as an arbitrage opportunity (we also used $0, $100, and $1,000 for robustness and achieved similar results). For example, a 2002 Dodge Neon with *Mileage* = 18,932 and *Valuation* = $6,468 was purchased (*Price*=$5,200) at the Denver facility on October 9, 2003. We identified five matching vehicles (i.e., 2002 Dodge Neons with *Mileage* between 18,000 and 19,000 and *Valuation* between $6,000 and $7,000) sold on October 14 at the Phoenix facility (mean *Price*=$6,660) and three matching vehicles sold on October 14 at the Dallas facility (mean *Price*=$5,433). Given the price differences, estimated transport costs, and transaction fees, we determined that Phoenix represented a profitable arbitrage opportunity but that Dallas did not. Thus, for the 2002 Dodge Neon purchased at Denver on October 9, we identified one arbitrage opportunity (in Phoenix).[7]

Panel A of Figure 2.3 shows that the number of arbitrage opportunities decreased by 72% over our sample period, from an estimated 1,532,232 in 2003 to 426,122 in 2010. Panel B shows that the number

---

6 See the appendix for a description of how we estimated transport costs. Other potential transaction costs involved in spatial arbitrage include taxes and the cost of capital. Taxes are not relevant because dealers do not pay taxes when purchasing vehicles in the wholesale market (tax is collected on retail transactions). The cost of capital is relevant if arbitrageurs purchase vehicles using debt (e.g., a line of credit) and must pay interest until they retire the debt. This cost is negligible for our analysis because the arbitrageurs hold vehicles for a very short time (no more than 7 days in our focal analysis); i.e., there is little time for interest to accrue.

7 This approach assumes that moving a matched vehicle from the source facility to the destination facility would not change the estimated price at the destination facility. This is questionable, because the additional supply at the destination facility would likely lower prices. This will cause our estimates of the number of arbitrage opportunities to be biased upward. However, this bias will be consistent across all 8 years of our sample. Given this consistency, the year-over-year decline shown in Figure 2.3 can still be interpreted as a decline in the number of arbitrage opportunities. Any other form of mismeasurement that exists across years will also not affect our conclusion.

of arbitrage opportunities that were exploited (i.e., the number of arbitrage transactions) also decreased over time (by 46%).[8] Panel D shows that electronic trading via both channels increased over time, with most of the electronic trading occurring via the webcast channel. Taken together, this indicates that efficiency increased as electronic trading became more prevalent. The correlations between the time series for the percentage of electronic trading and a) the number of arbitrage opportunities, and b) the number of arbitrage opportunities that were exploited (i.e., the number of arbitrage transactions) are -0.98 ($p<0.01$) and -0.95 ($p<0.01$).[9] This supports H1a. Panel C shows that the *percentage* of arbitrage opportunities that were exploited increased by 93% over our sample period, from 2.4% to 4.7%. The correlation between the time series for the percentage of electronic trading and the percentage of arbitrage opportunities that were exploited is 0.96 ($p<0.01$). This supports H1b and suggests that electronic commerce improved the ability for arbitrageurs to exploit inefficiencies, even as it decreased the overall number of inefficiencies. In other words, the market became more efficient, and the inefficiencies that persisted became easier for arbitrageurs to exploit.[10]

---

8 Overall, spatial arbitrage is rare. Less than 1% of purchased vehicles are later spatially arbitraged (with α=7). This is consistent with theoretical work about arbitrage (Shleifer and Vishny 1997).

9 To verify that the declines shown in Panels A and B of Figure 2.3 are not simply artifacts of a reduction in overall transaction volume, we also calculated the time series for the percentages of arbitrage opportunities and arbitrage transactions relative to total market transactions. These time series mirrored those shown in Panels A and B (the correlations between these time series and those in panels A and B are 0.99 and 0.98, respectively).

10 The reduction in arbitrage transactions (Panel B) suggests that the (negative) “opportunity reduction” effect of electronic commerce dominates the (positive) “opportunity exploitation” effect overall.

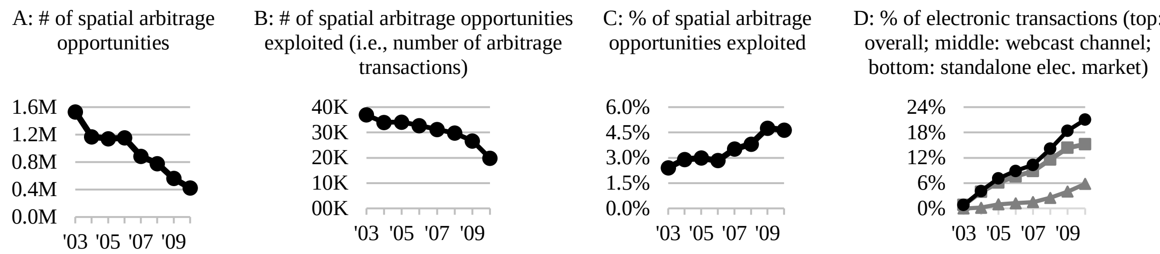


Figure 2.3: Spatial arbitrage and electronic trading trends over time.

2.5.2 Testing H2: The effect of different types of electronic commerce on the prevalence of spatial arbitrage

2.5.2.1 Testing the effect of the webcast channel: We tested the effect of the webcast channel on spatial arbitrage by leveraging the phased adoption of the webcast channel to conduct a quasi-natural experiment (e.g., Jensen 2007, Overby and Forman 2014). Because the webcast technology was deployed at different times for each lane at each facility, we observe many instances in which similar vehicles were sold at the same facility on the same day, some in lanes that were webcast enabled and some in lanes that weren't. We considered the former to be potential "treated" vehicles and the latter potential "control" vehicles. If whether a vehicle was sold in a webcast enabled lane was randomly assigned, then we could identify the treatment effect of the webcast channel via a simple comparison of outcomes for the treated and control vehicles. However, as with much observational data, that is not the case. When the webcast channel was deployed, sellers and facility managers (who collectively determine the lanes in which vehicles are offered) tended to use webcast enabled lanes for vehicles whose *Year/Make/Model* (e.g., 2002 Audi TT) were only available in a few facilities (so buyers from other areas could bid for them electronically)[11]

---

11 As support for this, we calculated the following for the first quarter of 2003: a) the number of vehicles of each *Year/Make/Model* sold in webcast-enables lanes (μ=5.4, σ=42.1), b) the number of vehicles of each *Year/Make/Model* sold overall (μ=156.4, σ=601.8), and c) the number of facilities at which vehicles of each *Year/Make/Model* were sold (μ=22.7, σ=23.4). We regressed (a) on (b) and (c). The coefficient for (c) was -0.09 (t=-4.90), such that a one standard deviation increase in the number of facilities at which a vehicle was sold was associated with a 39% decrease in the number of vehicles sold in webcast-enables lanes. We are also aware of the procedure by which vehicles were assigned to webcast enabled lanes

and that were relatively new with low mileage (so personal inspection of the vehicle was not required to assess quality). These assignment criteria became less relevant over time as the webcast channel was increasingly deployed. Nevertheless, vehicles sold in webcast and non-webcast enabled lanes differed in *VehicleYear*, *Make*, *Model*, *Mileage*, and *Valuation* (see Table 2.4).

Although assignment to webcast enabled lanes was not random, because we know the factors that influenced assignment (*VehicleYear*, *Make*, *Model*, and *Mileage*), we control for them. This allows us to conduct a meaningful comparison of treated and control vehicles. To do this (and to control for other factors), we matched control vehicles to treated vehicles on *VehicleYear*, *Make*, *Model*, *Mileage*, *Valuation*, *FacilityID*, *SellerType*, *SellerPctArbitraged*, and *SaleDate* (see Table 2.3). This increases the likelihood that the only material difference between control and treated vehicles are that the latter were sold on webcast enabled lanes. This allows us to attribute any significant differences in whether these vehicles are later arbitraged to the "treatment effect" of being sold in a webcast enabled lane. Essentially, the matched control vehicles serve as counterfactuals for what would have happened to the treated vehicles if they had not been treated, thereby allowing us to estimate a causal treatment effect (see Iacus et al. 2011). Matching on *FacilityID* is a key part of our identification strategy. In many cases, this allows us to compare vehicles sold in a webcast enabled lanes to very similar vehicles sold in a non-enabled lane just a few feet away. Matching on *SellerPctArbitraged* is also important, because this controls for a host of unobserved seller characteristics that influence the probability that vehicles they sell are arbitraged. One such characteristic is how effective sellers are at distributing vehicles to the "right" locations so that they don't sell for below-market discounts (and hence become good candidates for arbitrage).

We matched vehicles using exact matching and coarsened exact matching ("CEM"). Each treated

because one of the authors consulted with the intermediary on the initial implementation of the webcast channel.

vehicle could only be matched to a control vehicle with the same *FacilityID*, *VehicleYear*, *Make*, *Model*, and *SellerType*. We also restricted matches to vehicles sold in the same week. We coarsened *SellerPctArbitraged* into bins of width 0.1[12] and *Valuation* and *Mileage* into bins of width 1,000, and we only allowed matches between vehicles in the same bins.[13] Because there were essentially no potential control vehicles available after 2007 (because the webcast channel was almost fully deployed by then; see Figure 2.2), we restricted the analysis to observations between 2003 and 2007. The matching procedure yielded 19,095 matched strata/cells that each contained at least one treated vehicle and at least one control vehicle. For example, the procedure produced a stratum/cell for 2003 Ford Taurus's with *Mileage* between 22,000 and 23,000 and *Valuation* between \$10,000 and \$11,000 sold at the Minneapolis facility by institutional sellers whose *SellerPctArbitraged* was between 0.0036 and 0.004 in the second week of 2004; there was one treated vehicle and one control vehicle in this stratum.

To ensure that the procedure resulted in comparable matches, we examined the balance between the treated and control observations as follows. First, we set *DayOfWeek* = 1…7 based on which day of the week a vehicle was sold, with Monday = 1. Second, we calculated the means of *Valuation*, *Mileage*, *SellerPctArbitraged*, and *DayOfWeek* for the treated and control vehicles in each of the 19,095 strata in the matched sample, referred to as the *strata means*. We then used a t-test to examine whether these strata means differed significantly between the treated and control groups. As shown in Table 2.4, only

12 Because *SellerPctArbitraged* $\leq$ 0.004 for over half of the observations, we further coarsened the 0 to 0.1 bin into the following bins: 0 to 0.0003, 0.0003 to 0.0006, 0.006 to 0.009, …, 0.0033 to 0.0036, 0.0036 to 0.004, and 0.004 to 0.1. We also used a single bin for values between 0.5 and 1, given the rarity of observations with these values.

13 Coarsened exact matching temporarily coarsens each chosen variable into bins and exact matches on those bins, yielding a matched sample of control and treated observations. CEM then restores the original (non-coarsened) values of the variables for analysis. The CEM matching procedure may match an uneven number of control observations to treated observations. To account for this, CEM generates weights. Using these weights in an analysis procedure (such as regression) generates the sample average treatment effect. See Iacus et. al (2011) for details.

SellerPctArbitraged showed a significant differences between the two groups after matching, and this difference was of minimal practical significance (recall that matches are exact for all other variables). Table 2.4 also compares the variables across groups before matching.[14] Overall, we believe that our matches are precise enough to satisfy the un-confoundedness condition (aka, selection on observables) for valid matching estimation (Imbens, 2004). However, despite this precision, it is possible that *unobserved* differences could make the control vehicles inappropriate counterfactuals for the treated vehicles. For this to be a problem, the following conditions would have to hold. First, there would have to be unobserved vehicle characteristics (not captured in our matching procedure) that are correlated with a vehicle being arbitraged. Second, sellers and managers at the facilities – who collectively determine the lane in which vehicles are offered – would have to know what these characteristics are. Third, sellers and facility managers would have to consistently offer vehicles with these characteristics in non-webcast enabled lanes, while offering the other vehicles in enabled lanes (or vice versa). Although we cannot be sure that these conditions do not hold, they are unlikely. For one thing, it is unlikely that sellers and facility managers would be able to identify the variables – beyond those included in our matching – that consistently predict arbitrage, partly because arbitrage occurs rarely. Also, sellers might want to predict arbitrage if by doing so they could identify "mis"-distributed vehicles and move them to a more advantageous selling location, thereby retaining the arbitrageurs' profits for themselves. However, Overby and Clarke (2012) showed that sellers have little incentive to try to do this, because the revenues they forgo when they "mis"-distribute a vehicle that is later arbitraged are minimal compared to their total revenues. Last, there may be unobserved variables (e.g., scratches, dents) that influence both whether a

14 Although the means of the variables in the matched sample differ from those in the full data, we believe the matched sample is a good representation of the overall data. This is because the 95% confidence intervals around *Mileage* and *Valuation* in the matched sample cover 94% of the transactions in our data. I.e., the matched sample contains many matches of not only low mileage, high value vehicles but also of high mileage, low value vehicles.

vehicle is offered in a webcast enabled lane and its price. But if such variables cause a price discount (premium) at an arbitrageur's source location, they will also cause a discount (premium) at the destination location, such that potential arbitrage profits would be unaffected. Thus, such variables should not affect the likelihood of arbitrage. We also considered whether buyer heterogeneity across the webcast enabled and non-webcast enabled lanes might confound our result. We found this to be unlikely, because almost all buyers who purchased in webcast lanes also purchased in non-webcast enabled lanes, and vice versa (see the appendix for details). In addition, we conducted a sensitivity analysis (Rosenbaum 2002) to assess how influential any unobservables would have to be to alter our conclusion (discussed below).

Table 2.4: Descriptive statistics for treated and control observations for testing the effect of the webcast channel, before and after matching. Observations from 2003 to 2007.

Variable	n: strata	n: treated vehicles	n: control vehicles	Mean: treated vehicles	Mean: control vehicles	Difference in means (t-stat)
A: Before Matching						
Valuation	n/a	15,892,484	9,058,903	11,841.36	8,521.38	3,319.98 (1049.53)
Mileage	n/a	15,892,484	9,058,903	54,291.36	76,537.35	-22,245.99 (-1047.43)
SellerPctArbitraged	n/a	15,892,484	9,058,903	0.0068	0.0067	0.0001 (5.12)
DayOfWeek	n/a	15,892,484	9,058,903	3.12	3.04	0.08 (201.49)
B: After Matching						
Valuation	19,095	36,027	26,810	11,960.21	11,960.52	-0.31 (-0.25)
Mileage	19,095	36,027	26,810	38,633.18	38,636.40	-3.22 (1.26)
SellerPctArbitraged	19,095	36,027	26,810	0.00538	0.00530	0.00008 (2.01)
DayOfWeek	19,095	36,027	26,810	3.13	3.14	-0.01 (1.57)

Using the matched sample, we fitted the following logistic regression model to test the treatment effect of webcast enablement on whether a vehicle is arbitraged: $logit(probability(Arbitraged_j = 1)) = \beta_0 + \beta_1 * WebcastEnabled_j + \varepsilon_j$. $Arbitraged_j$ is set to 1 if vehicle *j* was arbitraged and 0 otherwise. $WebcastEnabled_j$ is set to 1 if vehicle *j* was sold in a webcast enabled lane and 0 otherwise. We fitted the model on the matched sample using weighted regression, with the weights provided by the CEM procedure (see footnote 13.) We set α=7 to delineate whether a vehicle was arbitraged in our focal model, and we varied this for robustness. We also fitted the model using two alternative specifications: a) a rare events logistic regression model, and b) a linear probability model of the form $Arbitraged_j = \beta_0 + \beta_1 * WebcastEnabled_j + \varepsilon_j$. Results are virtually identical to those we report. In other unreported analysis, we estimated the model after adding $Valuation_j$, $Mileage_j$, $SellerPctArbitraged_j$, $SaleDate_j$, and indicator variables for each $FacilityID_j$ as explanatory variables. These variables are already accounted for via the matching procedure, and their inclusion has no substantive effect on the $WebcastEnabled_j$ coefficient. ($SellerPctArbitraged_j$ is positive and significant in this specification, as expected.)

Table 2.5: Treatment effects of the vehicle being purchased on a webcast enabled lane (panel A) and in the standalone electronic channel (panel B) on whether the vehicle is later arbitraged.

A: Webcast Channel	
WebcastEnabled$_j$ (β_1)	-0.351 (0.137) ***
Intercept (β_0)	-5.494 (0.096) ***
n	62,837
Log likelihood; χ^2 (1)	-1425; 6.55 ***
B: Standalone Electronic Market	
StandaloneElectronicMarket$_j$ (β_1)	0.644 (0.102) ***
Intercept (β_0)	-5.596 (0.075) ***
n	83,611
Log likelihood; χ^2 (1)	-2416; 39.61 ***
The dependent variable is the probability that the vehicle is later arbitraged. Model estimated via logistic regression. Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Results shown with α=7, where α denotes the number of days between flips used to delineate spatial arbitrage.	

Panel A of Table 2.5 shows the results. The treatment effect of webcast enablement is captured by β_1. The effect is negative and significant; treated vehicles (i.e., those sold in webcast enabled lanes) were approximately 29%-39% less likely to be arbitraged than were control vehicles, depending on the value of α.15 This indicates that the webcast channel reduces the prevalence of spatial arbitrage. We also tested the effect of the webcast channel using McNemar's test (1947). Because this test requires matched pairs, we restricted the matched sample to only those strata/cells that contained one treated vehicle and one

15 We also used α=5, 14, and 28 days. We obtained the 29%-39% estimates by exponentiating the β_1 coefficients and by analyzing the results of the linear probability model (we divided β_1 by β_0 from the linear probability model.)

control vehicle (n=11,873 strata). For each of the matched pairs, one of four outcomes is possible: a) both the control and the treated vehicle were arbitraged (n=1 in our case), b) only the control vehicle was arbitraged (n=77), c) only the treated vehicle was arbitraged (n=52), and d) neither vehicle was arbitraged (n=11,743). McNemar's test examines whether the probabilities for the (b) and (c) outcomes are statistically different. As above, treated vehicles were significantly less likely to be arbitraged than were control vehicles ($\chi^2_{(1)}$ =4.84; $p<0.05$). The percentage decrease between (b) and (c) is approximately 32%, which matches the effect size reported above.

We examined how sensitive our results might be to the possibility that unobserved variables cause vehicles that are likely to be arbitraged to be offered in non-webcast enabled lanes. If that were the case, then the effect we observe might be attributable to these unobservables and not to the treatment effect of the webcast channel. We conducted a sensitivity analysis based on McNemar's test to see how influential these unobservables would have to be to alter our conclusion (see Rosenbaum 2002, §4.3.2 for details on this procedure). We concluded that in order to attribute the higher rate of arbitrage in non-webcast enabled lanes to unobservables, the unobservables would need to: a) be a near perfect predictor of arbitrage, and b) produce a 17% increase in the odds of a vehicle being offered in a non-webcast enabled lane (i.e., Γ=1.17 in sensitivity analysis notation). Although there is no consensus about the appropriate size for Γ in social science research, Γ=1.2 is around average (Sen, 2014).16

2.5.2.2 Testing the effect of the standalone electronic market: To examine the effect of the standalone electronic market on spatial arbitrage, we used a similar matching procedure as above, with one major change. We considered vehicles sold in the standalone electronic market to be potential treated vehicles, with vehicles sold in the physical market – regardless of whether they were sold in a webcast

16 No value of Γ can prove that a matching procedure is or is not valid. Γ is simply an indication of how much of a confounding influence unobserved variables would have to have to alter a conclusion. For the reasons discussed above, we do not believe that unobserved variables confound our conclusion.

enabled lane – as potential control vehicles. We ran the matching procedure using the data from January 1, 2005 to December 31, 2010, as we observe minimal transaction volume in the standalone electronic market prior to 2005. The matching procedure yielded 26,937 matched strata consisting of 83,611 vehicles: 29,612 treated vehicles and 53,999 control vehicles. The balance between treated and control observations was good except for the *DayOfWeek* variable, as shown in Table 2.6. As a robustness check, we assessed whether the imbalance for *DayOfWeek* affected our results by exact matching on observations with the same *DayOfWeek*. These results are consistent with those we report below. Our regression specifications were identical to those for analyzing the effect of the webcast channel, except we replaced $WebcastEnabled_j$ with $StandaloneElectronicMarket_j$.

Table 2.6: Descriptive statistics for treated and control observations for testing the effect of the standalone electronic market, before and after matching. Observations from 2005 to 2010.

Variable	n: strata	n: treated vehicles	n: control vehicles	Mean: treated vehicles	Mean: control vehicles	Difference in means (t-stat)
A: Before Matching						
Valuation	n/a	723,637	27,478,654	17,324.16	10,715.69	6,608.46 (519.39)
Mileage	n/a	723,637	27,478,654	39,916.42	66,065.39	-26,148.97 (-423.81)
SellerPctArbitraged	n/a	723,637	27,478,654	0.0054	0.0062	-0.0007 (30.19)
DayOfWeek	n/a	723,637	27,478,654	2.82	3.09	-0.27 (-219.81)
B: After Matching						
Valuation	26,937	29,612	53,999	14,650.68	14,650.73	-0.04 (0.047)
Mileage	26,937	29,612	53,999	24,936.70	24,939.47	-2.76 (1.22)
SellerPctArbitraged	26,937	29,612	53,999	0.00560	0.00563	-0.00003 (-2.17)
DayOfWeek	26,937	29,612	53,999	3.05	2.70	0.35 (40.00)

Results of the logistic regression appear in Panel B of Table 2.5; results of the other models are virtually identical. β_1 is positive and significant. Vehicles purchased in the standalone electronic market are 55%-117% more likely to be arbitraged than vehicles in the physical market, depending on the value of α. We also tested the effect of the standalone electronic market using McNemar's test, as above. In this test, the "both arbitraged" outcome has n=2 observations, the "control arbitraged, treated not" outcome has n=64, the "treated arbitraged, control not" outcome has n=117, and the "both not" outcome has

n=15,906. Treated vehicles were significantly more likely to be arbitraged than were control vehicles ($\chi^2_{(1)}$ =15.52; p<0.01); the percentage increase is approximately 83%.

Sellers decide whether to offer vehicles in the standalone electronic market or the physical market, and they also decide what prices to accept. If unobserved factors influence these decisions and whether a vehicle is likely to be arbitraged, then our result could be biased. One possibility is that sellers accept below-market prices for vehicles in the standalone electronic market (i.e., they "dump" vehicles for less than they are worth), which would create arbitrage opportunities. This does not appear to be the case, because the *Price/Valuation* ratio is significantly higher for vehicles sold in the standalone electronic market than in the physical market (101% to 99%; p<0.01). Another possibility is that sellers offer in the standalone electronic market vehicles with unobserved vehicle characteristics that are correlated with arbitrage. For this to be true, sellers would have to: a) know which unobservables are correlated with arbitrage, b) identify vehicles that have these unobservables, and c) offer these vehicles consistently on the standalone electronic market and not in the physical market. For the reasons described in §2.5.2.1, we believe this to be unlikely. In addition, we used the same procedure as above to analyze how sensitive our results are to the possibility that unobserved variables cause vehicles that are inherently likely to be arbitraged to be offered in the standalone electronic market. To attribute the higher rate of arbitrage in the standalone electronic market to these unobservables, they would need to predict arbitrage almost perfectly, and they would need to produce more than a 48% increase in the odds of a vehicle being offered in the standalone electronic market (Γ=1.49). Also, see the appendix for a discussion of why buyer heterogeneity is unlikely to confound our conclusion.

Overall, H2 is supported. The two forms of electronic commerce have different effects on the prevalence of spatial arbitrage, with the form of electronic commerce that supports transaction immediacy (the standalone electronic market) having a more positive effect.[17]

2.5.2.3 Analysis of the mechanisms behind the effects of the two electronic channels: According to our theory, the expanded reach that electronic commerce provides should reduce arbitrage opportunities by helping remotely-located buyers – who might otherwise be potential "downstream" customers for arbitrageurs – purchase vehicles directly from source locations. Our theory also suggests that expanded reach should help arbitrageurs identify and exploit previously hidden arbitrage opportunities. It also suggests that arbitrageurs should leverage transaction immediacy to quickly identify and exploit arbitrage opportunities originating in the standalone electronic market. Whether these mechanisms increase or decrease the amount of arbitrage depends on whether the "opportunity reduction" effect outpaces the "opportunity exploitation" effect.

To explore this for the webcast channel, we first confirmed that the webcast channel led to expanded buyer reach. We used the matched sample and the regression specifications from §2.5.2.1 to test the treatment effect of a vehicle being offered in a webcast enabled lane on the likelihood of its being purchased by a remotely-located buyer (*RemoteBuyer*). We set *RemoteBuyer* = 1 if the distance between the buyer and the facility at which the vehicle was located (i.e., *BuyerDistance*; see Table 2.3) was at least one standard deviation above the mean (and two standard deviations for robustness); *RemoteBuyer* = 0 otherwise. Vehicles purchased from webcast enabled lanes were 15%-35% more likely to be purchased by a remote buyer ($p<0.01$), depending on the measure of "remote".[18] We also tested the treatment effect of the webcast channel on vehicle price (*Price*), finding a 0.5% increase ($p<0.01$), perhaps due to an increase in the number of buyers bidding on the vehicles (although our data do not report the number of

17 The effects of the two channels are statistically different from each other because one is negative and significantly different from zero and the other is positive and significantly different from zero.

18 This corroborates a similar finding in Overby and Forman (2015).

bidders). The *RemoteBuyer* result could reflect both the "opportunity reduction" effect (due to regular buyers purchasing at remote locations) and the "opportunity exploitation" effect (due to arbitrageurs finding vehicles in remote locations that can be profitably arbitraged). But the *Price* result should pertain only to "opportunity reduction" because it increases the arbitrageurs' cost of sourcing vehicles. On balance, the "opportunity reduction" effect seems to dominate the "opportunity exploitation" effect, generating the negative effect of the webcast channel.

We used an analogous procedure to test the effects of the standalone electronic market on *RemoteBuyer* and *Price*. Results are similar; vehicles purchased in the standalone electronic market were 72%-83% ($p<0.01$) more likely to be purchased by a remote buyer (depending on the measure of "remote") and had 1.1% higher prices ($p<0.01$). The key difference between the webcast channel and the standalone electronic market is that the latter provides transaction immediacy, which our theory suggests will help arbitrageurs identify and exploit arbitrage opportunities. If this is happening, then we should see arbitrageurs purchasing undervalued vehicles in the standalone electronic market very soon after they are listed there. To explore this, we obtained supplemental data that contained additional variables for 55% (*n*=337,295) of the transactions that occurred in the standalone electronic market between 2007 and 2010.[19] We used two variables in particular: the date the vehicle was listed on the standalone electronic market (*ListingDate*) and whether a vehicle was purchased via auction (either before or after the reserve price was met) or via "Buy Now" (*TransactionType*). Using the supplemental data, we calculated *DaysToSale*, which is the number of days between a vehicle's *ListingDate* and *SaleDate*. We also

19 These additional variables were not available before 2007 or for all transactions between 2007 and 2010. We checked the representativeness of the transactions for which the variables were available by comparing vehicles' *Mileage*, *VehicleYear*, and *Valuation* between these supplemental data and the full set of transactions in the standalone electronic market between 2007 and 2010. The mean *Mileage* for the supplemental data (full data) was 37,457 (39,968), the mean *VehicleYear* was 2006.8 (2006.5), and the mean *Valuation* was 17,625 (17,186). Although these means are statistically different at $p<0.01$, the supplemental data appear reasonably representative.

calculated *PriceValRatio*, which is the ratio of a vehicle's *Price* to its *Valuation*. We computed the mean of these two variables, both in aggregate and by *TransactionType*. Results appear in column A of Table 7. Of the 337,295 vehicles purchased in the supplemental data, 1,967 of these were spatially arbitraged. Column B of Table 7 shows the mean of *DaysToSale* and *PriceValRatio* for the arbitraged vehicles.

Table 2.7: Statistics for DaysToSale *and* PriceValRatio *for transactions in the supplemental data.*

	A: Supplemental data: All transactions		B: Supplemental data: Arbitrage transactions	
	n	*Mean (St. Dev.)*	*n*	*Mean (St. Dev.)*
DaysToSale: All Transactions	337,295	1.76 (2.85)	1,967	0.81 (1.87)
- Buy Now Transactions	195,062	1.91 (3.36)	1,059	0.76 (2.13)
- Auction Transactions (above hidden reserve price)	119,732	1.69 (1.81)	690	0.96 (1.57)
- Auction Transactions (below hidden reserve price)	22,501	0.85 (2.25)	218	0.61 (1.27)
PriceValRatio: All Transactions	337,295	1.01 (0.17)	1,967	0.95 (0.13)
- Buy Now Transactions	195,062	1.02 (0.18)	1,059	0.96 (0.14)
- Auction Transactions (above hidden reserve price)	119,732	1.01 (0.15)	690	0.96 (0.11)
- Auction Transactions (below hidden reserve price)	22,500	0.97 (0.14)	218	0.92 (0.10)

Consistent with our theorizing, arbitrageurs used the standalone electronic market to purchase undervalued vehicles more successfully than did "regular" buyers: the mean *PriceValRatio* paid by arbitrageurs was 95%, whereas the overall mean was 101% (the difference between these means is significant at $p<0.01$). Also, arbitrageurs purchased vehicles very soon after they were listed: arbitrageurs waited only 0.81 days (on average), which is less than half as long as the overall average of 1.76 days (the difference is significant at $p<0.01$). The time between listing and purchase for arbitrageurs was particularly short for Buy Now transactions and for auction transactions that occurred before the reserve price was met. This suggests that arbitrageurs likely use the standalone electronic market to scan for vehicles with undervalued Buy Now prices, leveraging the market's transaction immediacy to purchase them quickly. Also, arbitrageurs likely are quick to register "low-ball" bids for vehicles that – if accepted – lead to profitable arbitrage opportunities. Although many of these bids are likely beaten, some sellers accept them, perhaps because they represent a quick way to sell a vehicle. Neither of these behaviors is available to an arbitrageur when sourcing vehicles in the physical market, regardless of whether he is participating via webcast or physically. We explored this further by estimating the profit for each arbitrage transaction using the formula noted in §5.1. The average arbitrage profit is $781 (st. dev. 1,030) when the vehicle was sourced from the standalone electronic market and $672 (st. dev. 710) when the vehicle was sourced from the physical market (via either the traditional physical channel or the webcast channel). This difference in profits is significant at $p<0.01$. Also, the percentage of arbitrage transactions that were profitable is higher for vehicles sourced in the standalone electronic market (91.6% vs. 88.6%; $p < 0.01$). The overall pattern of results suggests that the arbitrage "opportunity exploitation" effect is much stronger in the standalone electronic market than in the webcast channel because the former supports transaction immediacy. As a result, the "opportunity exploitation" effect appears to outpace the "opportunity reduction" effect in the standalone electronic market, generating the positive effect of the standalone electronic market on spatial arbitrage.

2.5.3 Testing H3 and H4: Factors that affect arbitrageur behavior of where to source products

We used a discrete choice model to study how arbitrageurs choose where to source vehicles. Fitting a choice model requires the researcher to define the set of alternatives available to the decision-maker (referred to as the *choice set*) and to specify a utility function for each alternative (Train 2009). We observe the facility at which arbitrageur i on day t sourced a vehicle(s) that he later arbitraged; this is the "chosen" alternative in each choice set. We defined the "non-chosen" alternatives in the choice set as those facilities other than the chosen facility: a) that were open on day t, and b) at which arbitrageur i made a purchase during the sample period. We modeled the utility of each facility k to arbitrageur i at time t as $U_{ikt} = \beta_{0,k} + \beta_1 * PctOfferedWebcast_{kt} + \beta_2 * Distance_{ik} + \beta_3 * Distance_{ik} * NearbyFacilities_k + \beta_4 * Supply_{kt} + \beta_5 * Supply^2_{kt} + \beta_6 * PctSold_{k(t\text{-}30)} + \beta_7 * PctSoldLowPrice_{k(t\text{-}30)} + \beta_8 * GeoPriceDispersion_{k(t\text{-}30)} + \varepsilon_{ikt}$. We describe the explanatory variables in the utility function in Table 8.

Table 2.8: Variables used in the discrete choice model of where arbitrageurs source vehicles.

Variable	Description	Mean (St. Dev.)
PctOffered Webcast$_{kt}$	The percentage of vehicles at facility *k* that were offered in webcast enabled lanes on day *t*.	0.62 (0.41)
Distance$_{ik}$	The distance in miles between the zip codes of arbitrageur *i* and facility *k*.	535.23 (528.18)
NearbyFacilities$_{k}$[a]	The number of facilities within 350 miles of facility *k*.	20 (8)
Supply$_{kt}$	The number of vehicles offered at facility *k* on day *t*.	783.40 (765.07)
PctSold$_{k(t-30)}$	The percentage of vehicles offered at facility *k* in the 30 days prior to day *t* that were sold.	0.61 (0.12)
PctSoldLow Price$_{k(t-30)}$[b]	The percentage of vehicles sold at facility *k* in the 30 days prior to day *t* that sold for less than 90% of their *Valuation*.	0.20 (0.10)
GeoPrice Dispersion$_{k(t-30)}$	The average geographic price dispersion of the vehicles offered at facility *k* in the 30 days prior to day *t*. We measured this as follows. First, for each vehicle offered at facility *k* on day *t*, we created a list of all facilities at which vehicles of the same year/model were purchased in the 30 days prior to day *t*. Second, we calculated the average price for those vehicles at each facility during that time period. Third, we took the standard deviation of these average prices across facilities. This gave us a measure of how much the price of each vehicle offered at facility *k* on day *t* varied across facilities. Fourth, we averaged these measures for all vehicles offered at facility *k* on day *t*.	1537.77 (242.00)

[a] We also ran the model using the following different thresholds to define *NearbyFacilities$_k$*: 100 miles, 200 miles, and 700 miles. Results are substantively unchanged.
[b] We also ran the model using the following different thresholds to define *PctSoldLowPrice$_{k(t-30)}$*: 80% and 85%. Results are substantively unchanged.

We included *PctSoldLowPrice$_{k(t-30)}$* and *GeoPriceDispersion$_{k(t-30)}$* to test arbitrageur preferences for facilities at which a high percentage of vehicles are likely to be available for low prices and at which prices for the vehicles offered vary widely across facilities (H3a and H3b). We included *Distance$_{ik}$* to test arbitrageur preferences for facilities close to them (H4a). We tested H4b (that arbitrageurs prefer facilities that are relatively difficult for other traders to access) in two ways. First, we interacted *Distance$_{ik}$* with *NearbyFacilities$_k$*, reasoning that facilities located near other facilities are relatively easy for other traders to access, given that the location of the facilities closely matches the population density of the U.S.[20] Second, we included *PctOfferedWebcast$_{kt}$*, reasoning that the more the vehicles at a facility are available

20 We didn't include *NearbyFacilities$_k$* as a standalone variable because it is a constant for each facility *k* and cannot be estimated separately from each facility's alternative-specific constant, which is represented by $\beta_{0,k}$.

via webcast, the more accessible (electronically) that facility is to traders. Further, including $PctOfferedWebcast_{kt}$ allowed us to test whether arbitrageurs prefer sourcing vehicles at facilities at which the webcast channel has been deployed only minimally or not at all, which one might expect given the negative effect of the webcast channel on spatial arbitrage shown in §5.2.1. We included $Supply_{kt}$, $Supply^2_{kt}$, and $PctSold_{k(t-30)}$ to account for the role that facilities' size and liquidity play in arbitrageurs' sourcing decisions. We used 30-day lagged variables (noted by the *t-30* subscript) for $PctSold_{k(t-30)}$, $PctSoldLowPrice_{k(t-30)}$, and $GeoPriceDispersion_{k(t-30)}$ because the contemporaneous values of these variables are unknown to arbitrageurs when they choose the facility at which to purchase. We used contemporaneous variables for the other variables because arbitrageurs either already know them (e.g., $Distance_{ik}$) or can calculate them based on the "pre-sale" list of vehicles posted in advance on the intermediary's web site. We included alternative-specific constants to capture the latent utility (i.e., the fixed effect) of each facility *k* (represented as the $\beta_{0,k}$ term) and fitted the model using the multinomial logit specification. Results appear in Table 9.

Table 2.9: Results of the discrete choice model of where arbitrageurs source vehicles.

	Coefficient (Std. Error)
$PctOfferedWebcast_{kt}$ (β_1)	-0.3420 (0.0225) ***
$Distance_{ik}$ (β_2)	-0.5025 (0.0322) ***
$Distance_{ik}$*$NearbyFacilities_k$ (β_3)	-0.0383 (0.0017) ***
$Supply_{kt}$ (β_4)	3.5414 (0.0217) ***
$Supply^2_{kt}$ (β_5)	-0.5477 (0.0054) ***
$PctSold_{k(t-30)}$ (β_6)	-2.1817 (0.0606) ***
$PctSoldLowPrice_{k(t-30)}$ (β_7)	1.0042 (0.1072) ***
$GeoPriceDispersion_{k(t-30)}$ (β_8)	0.0041 (0.1030)
Alternative specific constants	included
Number of choices in choice set (Min, mean, max)	2, 5.9, 48
n (total number of choices)	698057
Log Likelihood	-115480
The number of days used to delineate spatial arbitrage (α) is set to 7 in this analysis. Results are consistent for α=5, α=14, and α=28. Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). To ensure that coefficients are of similar magnitude for reporting purposes we divided $Supply_{kt}$, $Distance_{ik}$ and $GeoPriceDispersion_{k(t-30)}$ by 1000.	

H3a is supported; the coefficient for $PctSoldLowPrice_{k(t-30)}$ is positive and significant. We used the model estimates to simulate the size (i.e., practical significance) of this effect as follows. We simulated the percentage change in the number of times arbitrageur *i* chose facility *k* when $PctSoldLowPrice_{k(t\text{-}30)}$ = 10% vs. 30% (i.e., one standard deviation below and above the mean). We did this for each facility. On average, this increased the probability of choosing a facility by approximately 20%. H3b is not supported; the coefficient for $GeoPriceDispersion_{k(t\text{-}30)}$ is positive (as posited) but insignificant. H4a is supported; the coefficient for $Distance_{ik}$ is negative and significant. We simulated the size of the $Distance_{ik}$ effect by estimating the percentage change in the number of times an arbitrageur chose facility *k* when $Distance_{ik}$ = 250 vs. when $Distance_{ik}$ = 750, with $NearbyFacilities_k$ set at its mean, which is 20. The increased distance reduced the probability of choosing a facility by 21%. H4b is also supported; the interaction between $Distance_{ik}$ and $NearbyFacilities_k$ is negative and significant. This shows that if two facilities are equidistant, the arbitrageur will prefer the one that is more isolated. To examine the effect of $NearbyFacilities_k$, we set $Distance_{ik}$ at its mean and ran the simulations with $NearbyFacilities_k$ set to 12 and 28. This increased density (and greater accessibility) reduced the probability of choosing the facility by 14%. The negative and significant coefficient for $PctOfferedWebcast_{kt}$ also provides support for H4b, along with corroborating our earlier results about the effect of the webcast channel. To examine the size of this effect, we ran the simulations with $PctOfferedWebcast_{kt}$ = 24% and $PctOfferedWebcast_{kt}$ = 71%, which are the mean values for this variable in 2003 and 2004 (and represent close to a one standard deviation increase). This reduced the probability that an arbitrageur would choose the facility by 13%. $Supply_{kt}$ has a positive effect. This may be because a large supply increases an arbitrageur's chance of finding undervalued vehicles. This relationship is concave (i.e., $Supply^2_{kt}$ has a negative effect), but the inflection point does not occur until $Supply_{kt}$ = 2,532, which is more than 2 standard deviations above the mean (note that we scaled $Supply_{kt}$ by dividing by 1,000; see Table 2.9.) Arbitrageurs also prefer to source at facilities at which recent sales percentages have been low ($PctSold_{k(t\text{-}30)}$ is negative), i.e., those that have been relatively illiquid.

2.6 Conclusion

Markets can improve social welfare, but the degree to which they generate this benefit depends on the degree to which buyers and sellers match efficiently across geographic distance. Electronic commerce can help buyers and sellers match across distance by making it easier for them to find and transact with trading partners in remote locations. We examined the effect of two distinct forms of electronic commerce on market efficiency as measured by the prevalence of spatial arbitrage.

2.6.1 Contributions and summary of findings

We make three main contributions. Our first contribution is that we measure market efficiency via the prevalence of spatial arbitrage rather than via price dispersion. The prevalence of spatial arbitrage has several measurement advantages over price dispersion. First, it inherently accounts for the transaction costs of moving products between locations. If these transaction costs are lower than the price difference between locations (i.e., if supply and demand are inefficiently distributed), then arbitrage will occur. If not, it won't. Second, the spatial arbitrage measure accounts for unobserved product heterogeneity because the same product is traded at both the source and destination locations. Unobserved product heterogeneity can confound the price dispersion measure if the products whose prices are being compared across locations differ due to unobserved quality differences. Third, spatial arbitrage is based on the micro-level behavior of the agents – the arbitrageurs – who are most aware of whether buyers and sellers are matching efficiently across geography. Thus, arbitrageur behavior provides a direct window into a market's level of efficiency. A potential drawback to the spatial arbitrage measure is that spatial arbitrage transactions are difficult to observe. Using spatial arbitrage as the measure requires unique (and consistent) identifiers for individual products, traders, and market locations. Although such data are relatively elusive, they are becoming more common as trading activity increasingly moves online and item-level tracking becomes more widely adopted.[21] We hope that our study encourages others to adopt the prevalence of spatial arbitrage to measure market efficiency.

Our second contribution is that we study two distinct forms of electronic commerce, which helps us understand the theoretical mechanisms through which electronic commerce affects market efficiency. Both forms – the webcast channel and the standalone electronic market – provide expanded reach. This should reduce arbitrage opportunities by helping regular buyers purchase directly from source locations (thereby disintermediating arbitrageurs) while simultaneously improving arbitrageurs' ability to find and exploit arbitrage opportunities that s/he might otherwise miss. But only the standalone electronic market provides transaction immediacy, which should further enhance arbitrageurs' ability to find and exploit arbitrage opportunities before they dissipate. Our results reveal this nuanced effect: electronic commerce reduces arbitrage opportunities but improves arbitrageurs' ability to find and exploit those that remain. The "opportunity reduction" effect dominates the "opportunity exploitation" effect for the webcast channel, but the converse is true for the standalone electronic market, which we attribute to the transaction immediacy it provides. Thus, the net effect of the webcast channel on the prevalence of spatial arbitrage is negative, while the net effect for the standalone electronic market is positive. We summarize and illustrate the effects of the two channels in Table 2.10. The cumulative effect of both channels is negative, likely because the webcast channel was more widely used during the sample period.

21 An example of an item-level tracking system from the agricultural industry is the National Livestock Identification System in Australia (www.dpi.nsw.gov.au/agriculture/livestock/nlis).

Table 2.10: Summary of how the two electronic channels affect spatial arbitrage.

	Webcast Channel	Standalone Electronic Market
Effect of expanded purchasing reach on spatial arbitrage for:		
"Regular" buyers	Helps them purchase directly from "source" locations, reducing arbitrage opportunities. ↓	Helps them purchase directly from "source" locations, reducing arbitrage opportunities. ↓
Arbitrageurs	Helps them identify and exploit otherwise hidden opportunities in remote locations. ↑	Helps them identify and exploit otherwise hidden opportunities in remote locations. ↑
Effect of transaction immediacy on spatial arbitrage for:		
"Regular" buyers	(The webcast channel does not provide transaction immediacy.)	-- -
Arbitrageurs	(The webcast channel does not provide transaction immediacy.)	Helps them identify and purchase undervalued vehicles for later arbitrage. ↑
Overall Effect	<u>Negative</u>: "opportunity reduction" outweighs "opportunity exploitation". ↓	<u>Positive</u>: "opportunity exploitation" outweighs "opportunity reduction". ↑

Our third contribution is that we document several novel findings about arbitrage and how arbitrageurs behave. This is important because despite arbitrage's central place within economic theory, data limitations have made arbitrage transactions notoriously difficult to observe. As a result, a key mechanism in economic theory about efficient markets has been left largely unexamined; we have either taken it on faith or measured it indirectly. We overcome this by using highly granular data to measure spatial arbitrage at a transaction-level, which allows us to document new insights about how arbitrageurs behave. Among other findings, we find that arbitrageurs prefer to source vehicles at locations that are relatively difficult for other market traders to access (both physically and electronically), likely because these locations are isolated from the rest of the market.

Overall, we show that electronic commerce has a more nuanced effect on market efficiency than has previously been shown empirically (to our knowledge). Our results are consistent with prior research that electronic commerce improves efficiency by helping buyers and sellers trade across distance, thereby balancing supply/demand. But we also show that electronic commerce can improve efficiency by improving arbitrageurs' ability to identify and exploit remaining supply/demand imbalances, the act of which helps restore efficiency to a market.

In addition to these contributions to the academic literature, the study has several managerial implications. First, it is relevant for spatial arbitrageurs, because it illustrates how their business model is being impacted by the diffusion of electronic commerce. Although electronic commerce provides tools to make it easier for arbitrageurs to find and exploit arbitrage opportunities, it also reduces the number of opportunities. Thus, arbitrageurs should continuously increase the sophistication with which they identify and exploit market inefficiencies to maintain their profits. Second, the study has implications for sellers in spatially distributed markets who must choose where to sell their products. In inefficient markets – i.e., those in which supply and demand are imbalanced – this is a very important decision because prices may vary significantly across geography. However, as markets become more efficient through electronic commerce, these distribution decisions become less important, allowing sellers to allocate resources to other tasks. Third, the study has implications for market intermediaries who provide trading platforms. When spatial arbitrage occurs, two transactions are needed for a product to get from the seller at the source location to the buyer at the destination location. If the buyer and seller transact directly, then only one transaction is needed. Intermediaries who charge fees for each transaction might lose revenue as spatial arbitrage becomes less prevalent.

2.6.2 Limitations and future research

A limitation of our analysis is that it is specific to the wholesale used vehicle industry. However, well-functioning automotive markets are important in their own right, given the surprisingly large impact that the automotive industry has on the overall U.S. economy.[22] This importance is reflected in several academic studies focused on the industry (e.g., Dimoka et al. 2012). Testing whether the results hold for other industries represents an opportunity for future research. Another limitation of our analysis is that

---

22 In 2012, sales at car dealerships represented approximately 15% of total retail sales in the U.S., and dealership payroll represented approximately 12% of total retail payroll. Source: National Automobile Dealers Association (http://www.nada.org/Publications/NADADATA/).

many of the results are based on matching estimation in a quasi-natural experimental setting. We have matched on many important variables, including market facility, sale date, vehicle year, make, model, mileage, valuation, seller type, and the probability that a seller's vehicles are arbitraged. However, we cannot rule out the possibility that unobserved variables might confound our conclusions, although we have conducted sensitivity analyses to assess how influential these unobservables would need to be. Another limitation is that although we are able to observe spatial arbitrage transactions with a high level of precision, the precision is imperfect. It is possible that we have misclassified spatial arbitrage instances, although our results are robust to different measures. Last, space and scope limitations preclude us from examining other research questions, such as how arbitrageurs choose between the physical market and the standalone electronic market when sourcing vehicles for arbitrage and/or when selling vehicles to complete arbitrage transactions. This represents an opportunity for future research.

CHAPTER 3: SPATIAL ARBITRAGE AND ARBITRAGEUR SPECIALIZATION STRATEGIES

3.1 Introduction

Spatial Arbitrage[23] is a fundamental market feature contributing to efficient markets. In spatial arbitrage, a good purchased at one location is transported to another location by an arbitrageur and later sold. The arbitrageur hopes to earn a profit after the good is sold at the destination. If many instances of spatial arbitrage prevail in markets, it indicates inefficiencies in the market. However, over time, markets become more efficient as prices across geographic distances converge to a difference not exceeding the transaction cost (i.e., cost of transportation and associated fees), thereby reducing arbitrage profits to zero (Barrett 2008, Takayama and Judge 1964). When arbitrage profits are reduced to negligible levels in markets, arbitrageurs do not find it profitable to engage in arbitrage, and markets would be considered efficient. Though spatial arbitrage is fundamental to the study of markets, very few studies, given data limitations, have been able to identify and study spatial arbitrage. Even fewer are those studies that have attempted to study behaviors of arbitrageurs in the context of spatial arbitrage. Research identifying instances of spatial arbitrage and arbitrageur behaviors is rare because datasets with identifiers for the goods, traders, time, and locations are difficult to obtain. In this paper, using a unique dataset with micro-level identifiers and information pertaining to goods, traders, and locations, we study the behaviors of arbitrageurs and the effect of these behaviors on arbitrage profits.

Arbitrageurs play a key role in making markets more efficient and in reducing supply/demand imbalances. The primary motivation for arbitrageurs is economic profits (Shleifer and Vishny 1997). However, making risk-free profits in markets by means of spatial arbitrage is not easy. Scholars studying

[23] One of the fundamental concepts in finance is **arbitrage**, defined as “the simultaneous purchase and sale of the same, or essentially similar, security in two different markets for advantageously different prices” (Sharpe et. al ,1999)

behavioral finance have identified and theorized that capital, cognitive and, rational limitations of arbitrageurs inhibit arbitrage. Scholars have both empirically and analytically shown that these limitations play a key role in reducing the ability of arbitrageurs to exploit all available opportunities in markets (Barberis and Thaler 2003, Mullainathan and Thaler 2000). More recent research in behavioral finance—namely, the Adaptive Markets Hypothesis—has provided a framework for understanding these limitations faced by arbitrageurs, wherein market actors (i.e., both arbitrageurs and regular traders) alter their behavior as a reaction to environmental factors (Farmer 2002, Lo 2004).

The current paper's motivation stems from a need to understand the behaviors of arbitrageurs, the changes of these behaviors, and factors causing these changes (otherwise called adaptations) in markets. This paper has three goals. Firstly, we test and confirm the theoretical predictions pertaining to arbitrageur specialization as predicted by Shleifer and Vishny. Secondly, we group arbitrageurs based on patterns of specialization behaviors exhibited, thus confirming the predictions pertaining to the existence of species of traders. Finally, we examine the *why* by examining antecedents influencing these behavioral changes with respect to specialization.

In the current paper, we empirically examine the following research questions in the context of spatial arbitrage: *How does specialization affect arbitrage outcomes* (*i.e., economic profits and number of arbitrage instances*)*? How do specialization strategies of arbitrageurs evolve over time?*

There are three main reasons why these research questions pertaining to arbitrageur specialization are important. Firstly, spatial arbitrage in itself is a very important determinant of market efficiency and is an outcome of arbitrageurs behaving in a certain way. Understanding arbitrageur behaviors will help us understand why markets increase or decrease in efficiency over time. Secondly, theoretical predictions, with respect to arbitrageur specialization, have not, to our knowledge, been empirically tested for a lack of transaction-level data at the level of the arbitrageur. The dimensions along which arbitrageurs specialize and the extent to which their specialization influences profits is important to understand since these help inform long-held theoretical beliefs about arbitrageur behavior. Finally, the most recent contributions in behavioral finance that pertain to adaptive markets hypothesis have provided a framework to understand the evolutionary behaviors of all market actors. Testing the behavioral patterns of arbitrageurs in the context of spatial arbitrage will lend credibility to this recent theory. Understanding the antecedents influencing the evolution of specialization will provide a deeper understanding of arbitrageur behavior.

In this paper, we identify arbitrageur specialization choices and find that the type of asset and sourcing location are two important yet independent specialization choices that arbitrageurs make. Our results indicate that both asset specialization and source location specialization affect arbitrage profits and arbitrage intensity (i.e., the number of vehicles arbitraged in a given period). We identify unique groups of arbitrageurs based on how vehicle and location specialization behaviors evolve over time. We further analyze the antecedents that cause changes in specialization using a system of equations that can help identify the causal link among antecedents, specialization, and arbitrage outcomes (i.e., profits and arbitrage intensity as measured by the number of vehicles arbitraged). We find that intensity of arbitrage, available capital, and percentage of vehicles sourced from remote locations using electronic channels affect how arbitrageurs alter their specialization strategies.

The remaining sections of this paper are organized as follows: In section 3.2 we discuss prior literature. In section 3.3 we develop the theory and state our main hypotheses. In section 3.4 we describe the empirical context and test our main hypotheses. In section 3.5 we conclude and state limitations of this study.

3.2 Literature review

Our paper extends two streams of research: a) the literature on behavioral finance, the literature on limits of arbitrage, and the literature on adaptive markets hypothesis, and b) the literature on spatial arbitrage, electronic-commerce, and market efficiency.

3.2.1 Prior research on the limits of arbitrage

The efficient markets hypothesis stream of literature is one of the earliest streams of financial research and investment theory. The efficient markets hypothesis states that markets fully, accurately, and instantaneously incorporate all available information into market prices (Fama 1970, 1998). Though most research pertaining to efficient markets hypothesis focuses on financial instruments such as stocks, currency, or commodities, several scholars have looked at the effect of efficient markets hypothesis on products whose prices move based on available information. For example, prior research has tested the efficient markets hypothesis in the context of physical products in, for example, agricultural markets (Baffes 1991, Mackenzie and Holt 2002) and single family homes (Case and Shiller 1988). We develop the current paper in the context of the automobile market.

The efficient markets hypothesis research stream assumes that traders are rational and update their beliefs as and when new information becomes available in markets (Barberis and Thaler 2003). It is possible that traders do not update their beliefs correctly as market conditions change and that traders need not be rational. These non-ideal behaviors give rise to opportunities for arbitrage, which are exploited by those traders seeking profits and those who are willing to track market inefficiencies to the extent that the inefficiencies can be exploited for economic profit. The argument put forth by the proponents of the efficient market hypothesis is that when arbitrage opportunities surface in markets, a large number of small traders will exploit these opportunities, thereby quickly returning markets to their original equilibrium state.

The behavioral finance literature originating in a paper by Shliefer and Vishny (1998) called "The Limits of Arbitrage" shows that the argument put forth by efficient markets hypothesis is flawed. In this paper Shleifer and Vishny define arbitrage as a specialized activity that only a few market actors with sophisticated knowledge would conduct. The main assessment of scholars following this stream of research is that arbitrageurs are specialized, risk-averse traders whose main motivation is economic profits (Barberis and Thaler 2003, Shleifer 2000). In markets, arbitrage opportunities are plentiful and continue to exist despite the role of arbitrageurs (Shleifer 2000). However, as regular traders recognize these opportunities, they gradually buy into such, thus reducing opportunities for trade of arbitrageurs and increasing the risk for arbitrageurs. Arbitrageurs, being specialists in markets, also cannot alter their profit-making strategies fast enough to retain profitability. Capitalists (or investors) from whom arbitrageurs borrow capital withdraw capital available to arbitrageurs. This makes it difficult for arbitrageurs to further engage in trade when regular traders impinge on arbitrage opportunities. Thus, available capital for arbitrage determines the extent to which arbitrageurs are able to exploit opportunities in markets. We test capital availability in the current paper and study how the availability of capital affects both arbitrage profits and arbitrage intensity, which is a measure of how well arbitrageurs engage in markets.

Several papers have independently tested the predictions put forth by behavioral finance theory in the light of capital limitations, information limitations, and various types of risks faced by arbitrageurs. Some papers have identified and tested for different types of risks faced by arbitrageurs in financial markets (Gabaix et al. 2007, Mitchell and Pulvino 2001) and have ascertained that these risks indeed inhibit the ability of arbitrageurs. Similarly, scholars have tested and challenged capital limitations faced by arbitrageurs and have found that capital limitations, though exciting, can be avoided by various portfolio/trade choices made in the context of study (Hanson and Sunderam 2014, Hombert and Thesmar 2014, Ljungqvist and Qian 2014). Another stream of research shows how holding costs and transaction costs impede arbitrage (Pontiff 1996). As with the testing of efficient markets hypothesis in the context of product markets, spatial arbitrage has been related to behavioral finance and the limits of arbitrage literature by a few scholars, most notably those in the automobile sector (e.g., Overby and Clarke 2012).

A limitation faced by arbitrageurs that is rarely studied empirically–but that is often cited by many scholars—is that of specialization (Shleifer and Vishny 1997). While regular traders use portfolio diversification (Lintner 1965) for both long-term and short-term trades in order to reduce systemic and idiosyncratic risks to their profits, arbitrageurs prefer specialization as a strategy. There are two main reasons why specialization rather than diversification helps arbitrageurs. Firstly, specialization helps arbitrageurs gain and retain information with regards to profits earned while trading very specific assets. Given the fact that arbitrage is conducted within a short span of time (i.e., purchase and sale of the same asset almost instantaneously), profit uncertainty is reduced when arbitrageurs repeat the same pattern of purchase, especially when environmental conditions don't change (Kahneman and Tversky 1979, Mitchell and Pulvino 2001). Secondly, specialization helps an arbitrageur observe and differentiate his or her offerings from other arbitrageurs operating in markets, thereby avoiding competition and reducing risks to profits due to competition (Norton and Tenenbaum 1993).

One of the key disadvantages to specialization, especially when specialization behaviors don't change over time, is that specialization increases the arbitrageur's exposure to idiosyncratic risk. Idiosyncratic risk is a form of profit risk arising in markets due to reasons such as demand/supply shocks or sudden changes in information pertaining to trade, the cause of which remains unknown to the arbitrageur. Given that arbitrageurs are risk-averse and profit-seeking traders, specialization could have a negative effect when such shocks happen. For example, fluctuations in demand/supply when regular traders recognize easy opportunities for profits would cause arbitrageurs to exit (Gabaix et al. 2007). Another slow but deliberate change could be the adoption of e-commerce in traditionally physical markets. Electronic commerce enables buyers and sellers to trade directly, disintermediating the arbitrageur, as shown in Chapter 2. However, the rollout of new technology can also increase the arbitrageur's ability to identify newer arbitrage opportunities for sourcing, also demonstrated in Chapter 2.

Arbitrageurs, whose main motive is to retain risk-free profits, are known to reduce profit uncertainty by specialization. In the current paper, we study two key dimensions on which arbitrageurs specialize:

assets and sourcing locations. We further test theoretical predictions with respect to specialization and study how specialization affects arbitrage profits and arbitrage intensity.

3.2.2 Research on Adaptive Markets Hypothesis

A more recent stream of literature that attempts to bridge the gap between the efficient markets hypothesis and the limits of arbitrage is the Adaptive Markets Hypothesis (Lo 2004). The adaptive markets hypothesis attempts to reconcile theories on the limits of arbitrage and theories pertaining to the efficient markets hypothesis by providing a framework set in the light of the theory of evolution. In this paper, Andrew Lo states the following:

> *Efficient Markets Hypothesis supports market forces will always act to bring prices back to rational levels, implying that the impact of irrational behavior on financial markets is generally negligible, and therefore, irrelevant .But, this last conclusion relies on the assumption that market forces are sufficiently powerful to overcome any type of behavioral bias, or equivalently, that irrational beliefs are not so pervasive as to overwhelm the capacity of arbitrage capital dedicated to taking advantage of such irrationalities.. This is an empirical issue that cannot be settled theoretically, but must be tested through careful measurement and statistical analysis….*

Adaptive markets hypothesis relies on the fact that markets never reach a state of equilibrium and that profit opportunities will abound in markets, despite the role of arbitrageurs (Grossman and Stiglitz, 1980). When environmental changes occur either as a shock or gradually, arbitrage becomes risky. As a result, the arbitrageur is forced to alter behaviors in response to these environmental changes in order to remain profitable (Lo 2004, Lo 2005, Lo 2008)

Just as environmental factors play a role in shaping arbitrageur behavior, the arbitrageur's own personal preferences (called "behavioral biases") are shaped by the arbitrageur's assessment of market needs, personal preferences, geographic locations, and the desirability of a certain level of profits. These behavioral biases play an important role in determining how the arbitrageur responds to environmental changes in the market (Farmer 2002). The adaptive markets hypothesis states the following with respect to the evolution of traders"

> *While the former i.e. EMH may be viewed as the steady-state limit of a population with constant environmental conditions, the latter i.e. AMH involves specific adaptations of certain groups that may or may not persist, depending on the particular evolutionary paths that the economy experiences....*

This prediction sets the groundwork for testing the existence of subgroups of traders who exhibit common behavioral changes over time. Recent papers in finance have tested the adaptive markets hypothesis in the context of foreign exchange markets (Neely et al. 2006) and stock markets (Kim et al. 2011). However, these papers have used price movements to draw conclusions about behaviors and adaptations of traders. Price movements and variations are, in fact, byproducts of trade resulting from actions of regular traders and arbitrageurs. In the current paper, we use trader-level data along with asset-level data over several time periods to analyze the behaviors of arbitrageurs. We group arbitrageurs based on their specialization and identify different groups of arbitrageurs based on specialization patterns with respect to vehicle specialization and location specialization. Further, we study antecedents that affect the evolution of arbitrageur specialization strategy.

3.2.3 Literature in e-commerce, market efficiency, and spatial arbitrage

The literature on e-commerce and markets is a widely cited stream in information systems, economics, and management. Scholars have used various contexts to study the effect of electronic trade on market efficiency. For example, scholars have compared brick-and-mortar stores and electronic commerce stores (Brynjolfsson and Smith 2000). Similarly, scholars have studied the effects of the adoption of e-commerce channels (Bakos 1991, Jensen 2007) and the use of online trade channels

(Overby and Clarke 2012, Overby and Forman 2014) on market efficiency. One of the reasons cited for markets getting more efficient is that e-commerce has made it easier for both sellers and buyers to find and trade with each other by reducing search costs. Scholars have studied the role of arbitrageurs in making markets more efficient, especially in the presence of electronic trading channels (Overby and Clarke 2012). Chapter 2 shows that as electronic trading is adopted in markets, spatial arbitrage (and arbitrage opportunities) is reduced. However, the arbitrageur's ability to exploit opportunities increases over time. The effect of multiple electronic channels varies with respect to the channel features—namely, transaction immediacy and channel reach. Chapter 2 also shows that although overall arbitrage opportunities are reduced, arbitrageurs get more efficient at exploiting the remaining opportunities in markets. This idea is important to the current paper in which we study how sourcing via the electronic channels affects arbitrage profits. We also analyze how sourcing using electronic channels affects change in arbitrageur specialization behavior over time.

This paper's novel contributions are summarized below. Prior studies have used dual listed companies (De Jong, Rosenthal, and Van Dijk, 2009; Froot and Dabora, 1999), foreign exchange markets (Neely, Weller, and Ulrich, 2009), cross-listed shares (Gagnon and Karolyi, 2010), among others to indicate that arbitrageur limitations exist in markets. Each one of these studies has used price variations or inefficiencies exploited by arbitrageurs due to price variations in order to arrive at conclusions regarding arbitrage limitations. Price-based effects are, in fact, byproducts of trader behavior.

In the current paper, because of the richness of data, we are able to study arbitrageur behaviors directly using identifiers for the arbitrageurs and goods being traded. Additionally, this paper studies and recognizes the effect of specialization on arbitrage outcomes and the effect of specialization behavior of arbitrageurs, which, to our knowledge, has not been empirically validated in prior literature. Another unique contribution of our paper is that by directly observing arbitrage activity in markets and identifying each arbitrage transaction, we are able to classify different specialization behaviors exhibited by arbitrageurs in markets. This directly provides empirical evidence for the predictions in the adaptive markets hypothesis and shows how various antecedents affect the evolution of specialization behaviors in markets. These findings shed more light on arbitrage activity and empirically support the adaptive markets hypothesis.

3.3 Theory and Hypotheses

3.3.1 Arbitrage specialization

Professional arbitrage is a specialized and risky activity that involves considerable capital. Hence, only a few specialized traders engage in arbitrage. Over time, these traders learn to specialize on the asset type and sourcing locations and thus become more efficient (Gabaix et al., 2007; Shleifer and Vishny, 1997). Arbitrageur specialization over time occurs because of the need to reap higher profits by reducing profit uncertainty. Thus, we hypothesize the following:

H1: Increased asset specialization and location specialization is positively correlated with arbitrage profits.

3.3.2 Evolution of arbitrageur behavior: Effect of arbitrage intensity, e-commerce, and capital

Scholars who support the limits of arbitrage hypothesis claim that arbitrageurs are limited in their ability to exploit arbitrage opportunities. Different groups of arbitrageurs specialize differently in response to changes in markets due to behavioral biases and environmental factors (Lo, 2004). Arbitrageurs can specialize on assets and the locations from which they source assets. Eventually, economic profits earned by arbitrageurs determine whether they can continue to operate in markets over the long term since economic profits are the only motive for arbitrageurs in markets. As seen from H1 above, we see that increased specialization leads to increased profits, since arbitrageurs tend to reduce their risk. As environmental factors change, arbitrageurs have to alter their specialization by either sourcing their assets from different locations or by sourcing different types of assets in order to remain profitable. Environmental changes do affect existing strategies of arbitrageurs, causing them to change their specialization over time. Thus we hypothesize the following:

H2: Arbitrageur specialization strategies with respect to asset type and sourcing locations change over time.

Environmental changes such as the introduction of electronic commerce induce demand shocks, since they tend to increase the number of traders who can purchase the same good. This increase in demand affects equilibrium conditions in markets by moving prices away from their usual state, which the arbitrageur used to exploit for profits. As a result, arbitrageurs are forced to alter specialization strategies in response to such changes in order to remain profitable. As arbitrageurs increasingly engage in arbitrage in markets, they face various environmental changes, so they must evolve their strategies in order to attain a state of specialization, from which they can continue to trade with some level of certainty. More arbitrage leads to higher profits, provided that the specialization strategy is right.

Thus, we hypothesize the following:

H3: Increased arbitrage activity is positively correlated with the asset specialization and is positively correlated with location specialization: Increased arbitrage activity is positively correlated with profitability of the arbitrageur.

With electronic trading and the increased reach of electronic channels, arbitrageurs have increased access to asset types they specialize in from other locations. This increased reach, in some ways, offsets the reduction of opportunities caused by an increase in demand for the previously arbitraged good, as discussed above. From H1 above, we see that increased asset specialization is positively correlated with arbitrage profits. Thus, in order to offset the loss of arbitrage opportunities due to location specialization, arbitrageurs can increasingly source assets they specialize in from remote locations, given that they have increased reach from electronic trading. Thus, we hypothesize the following:

H4: Increased sourcing of vehicles from electronic trading channels is positively correlated with asset specialization and is negatively correlated with location specialization

One of the important limitations of arbitrage is the access to capital for arbitrage (Shleifer and Vishny, 1997). However, since an arbitrageur is a specialist by nature, increasing capital would mean that an arbitrageur can learn about newer vehicles without worrying about making profits. In other words, access to capital reduces the arbitrageur's risk. Arbitrageurs with increased capital would increase their participation in the market in the hopes of earning more arbitrage profits. Thus, the easiest way to increase market participation is getting to a state of higher specialization by increasing their arbitrage activity and finding the right strategy. Thus, increased capital causes an arbitrageur to be more specialized because it increases his or her market participation. Thus, we hypothesize the following:

H5: Increased capital availability is positively correlated with intensity of arbitrage and with both asset specialization and location specialization.

We proceed to test our hypothesis using the empirical context below.

3.4 Empirical Analysis

Our empirical context is the wholesale auto auction market. This market is geographically distributed across 80+ locations in the continental United States. An auction intermediary maintains these facilities. In this market buyers and sellers trade vehicles in auctions organized by the intermediary in the English auction format. Buyers are mostly wholesale automobile dealers. Sellers are either dealers or large enterprise traders, such as automobile rental companies or finance companies.

Each auction facility where vehicles are auctioned has multiple auction lanes numbered sequentially. On the day of sale, each vehicle is driven onto a particular lane, wherein an auctioneer solicits bids. The highest bidder wins the auction once the seller accepts the bidding price. Beginning in 2003, the auction intermediary started enabling remote auctions by installing video recording and voice equipment on physical lanes. This enabled remote buyers to submit bids from their computer terminals instead of physically being present at the facility. Also, the auction intermediary started selling vehicles using a standalone electronic market similar to eBay.com in 2005. Buyers could submit bids on vehicles and purchase vehicles through the website, and sellers could choose to list their vehicles on the website. Thus, there were mainly two modes of purchase: the physical mode where buyers and sellers were physically present at the auction and the electronic mode where the buyer purchased a vehicle either in competition with regular bidders through webcast or via the standalone electronic market. Our dataset has unique identifiers for buyers, sellers, goods traded, price, and sale date. The dataset is described in Chapter 2, Table 2.3. There were a total of 40,657,724 transactions in this market between 2003 and 2010 and 201,275 unique traders who purchase these vehicles between 2003 and 2010.

3.4.1 Spatial arbitrage

Following Overby and Clarke (2010), we define flips as those pairs of transactions in which the buyer i in the first transaction purchases a vehicle with VIN v at location j on date d and sells the same vehicle with VIN v at a location k on date e. We classify those flips as spatial arbitrage when the difference between the purchase date d and sale date e is $\alpha = 7$ days and the locations j and k are different.

3.4.1.2 Trader as an arbitrageur and delineating a subset of arbitrageurs for our study

Just as financial market traders use arbitrage as part of their profit-making strategy, arbitrageurs in the market for automobiles use spatial arbitrage as a profit making strategy. Thus, for most traders spatial arbitrage plays a dual role in this market. Firstly, they purchase vehicles for sale in their dealerships. Secondly, traders arbitrage a subset of vehicles. Arbitrage in itself involves considerable effort—though the returns are immediate and, if properly strategized, risk free. This is because arbitrageurs have to first purchase the vehicle at a source location by identifying the right vehicle at the right price at a t- α days, considering that they sell the vehicle on day t. Then, the arbitrageur has to transport the vehicle to another location within a short interval of time (e.g., possibly on the day of sale or the day following the sale). The transportation cost should be such that it doesn't reduce his or her profits too much. Finally, the arbitrageur has to sell the vehicle at another auction facility that might be far away. The arbitrageur also has to sell the vehicle at a price that gives him or her an acceptable level of profit. Each of these activities comes with a certain amount of uncertainty or unforeseen risk, which affects his or her overall profit risk. For example, a transportation delay would hold up capital and increase risk with respect to liquidity. Similarly, if the destination facility does not auction the vehicle within the given time, the arbitrageur would again increase the cost to capital by increasing the risk to his liquidity. If there is a sudden increase in supply of vehicles of the type the arbitrageur intends to sell at the location, the arbitrageur may be unable to sell the vehicle at a profitable price. Various types of uncertainties are introduced over the entire arbitrage transaction—the purchase of the vehicle, transportation, and the sale of the vehicle. The uncertainty introduced at each stage of the arbitrage process is in itself large enough to make spatial arbitrage unattractive for the regular dealer. The regular dealer is knowledgeable about his customer who comes to the dealer's lot to purchase a particular vehicle type, and most dealers would be satisfied by this opportunity. This is one of the reasons why only a few select traders engage in arbitrage in markets.

Next, we differentiate arbitrageurs from regular traders based on a few criterion, as stated below. Firstly, we have to differentiate arbitrageurs from regular traders because all arbitrageurs are also dealers. Arbitrageurs have two roles: that of a regular trader and that of an arbitrageur. Secondly, we have to identify arbitrage activity as sizeable enough to conduct meaningful econometric analysis with sufficient cross-sectional variation.

There are 13,195 traders who have flipped a total of 255,379 vehicles between 2003 and 2010 at least once. Amongst these 13,195 traders, none of them has arbitraged 100% of the vehicles purchased.

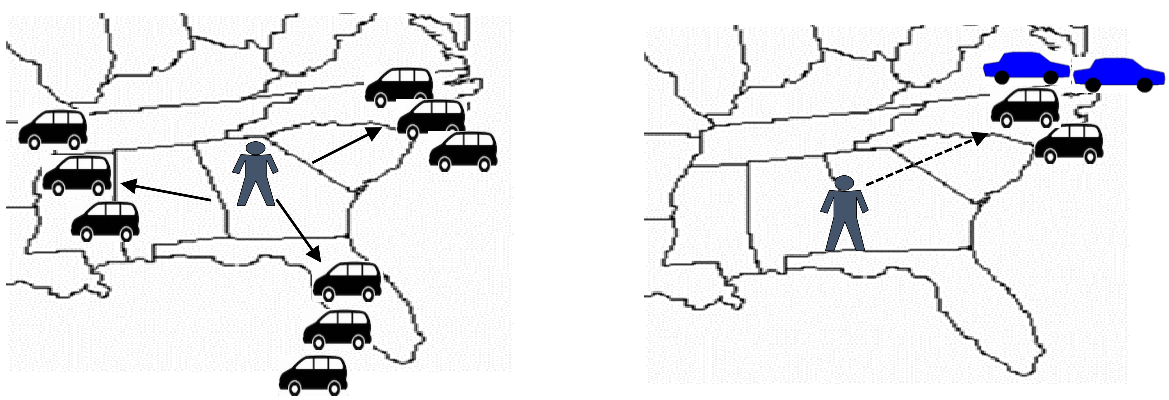
Thus, in order to differentiate a regular dealer from an arbitrageur, we do the following: Out of 13,195 traders, 5,602 have flipped only a single vehicle in the entire eight-year period; 11,107 traders have flipped between 1 and 10 vehicles in the entire eight-year period; and 12,475 traders have flipped between 10 and 50 vehicles in the entire period. In order to derive measures for vehicle specialization and source location specialization, we would need at least 3 vehicles per year from the sample. Furthermore, 720 of these traders have arbitraged at least 51 cars between 2003 and 2010, and we can reliably use these traders and their transactions as the focal dataset for our analysis. These sets of transactions constitute 1,95,634 arbitrage transactions or ~80% of all transactions we classify as spatial arbitrages in the market. This threshold is a reasonable assumption for studying arbitrageur behaviors for the following reasons. Firstly, as described above, spatial arbitrage involves considerable effort in order to be profitable. Arbitrageurs who arbitrage at least 50 vehicles would have either understood the nuances of arbitrage by handling all the uncertainties concerning purchase, transportation, and sale of the vehicle. They would either be engaging in arbitrage by successfully earning their desired level of profits or would have recognized that they are unable to make profits and would be in the process of winding up their arbitrage operations. Either way, arbitrageurs who are present in this subsample meet two conditions: that of deliberately engaging in arbitrage by specializing and that of being aware of market conditions affecting their profits.

3.4.2 Testing H1: The effect of specialization on arbitrage profits

3.4.2.1 Arbitrageur specialization

Prior literature in strategy, economics, and information systems has used statistical measures for indicating concentration. These concentration measures are used to indicate the size distribution of entities when multiple entities are present. For example, the market share of a particular firm's products amongst a group of firms that sell similar products is indicted by the Herfindahl-Hirschman Index (HHI). In other words, a concentration metric is a single real number that indicates how entities are distributed.

Prior research has shown that concentration plays an important role in various market studies (Curry and George, 1983). Research has shown that concentrations affect market power, business behavior, firm performance, and factors such as GDP, trade imbalance, etc., at the macro-level. In our research context, we use a Gini coefficient of the arbitrageur's preference for J. D. Power and Associates vehicle categorization[24] type to indicate vehicle specialization. We use a Gini coefficient to indicate an arbitrageur's preference for vehicle's make to test for robustness. Similarly, we use a Gini coefficient to indicate the location specialization, or the choice distribution of arbitrageurs with respect to source locations during a year. Both of these measures are calculated in the context of vehicles that are arbitraged in the market and meet our definitions for arbitrage. In addition to the Gini coefficient to indicate vehicle specialization and a Gini coefficient to indicate source location specialization, we also use the Theil index as an additional robustness metric. Figure 3.1(a) below illustrates the two strategies of arbitrageur profits. Arbitrageurs could specialize in sourcing locations and vehicle types, as illustrated below.



a. Arbitrageur specializes on vehicles but not on locations. He or she buys the same kind of vehicle from multiple locations.

b. Arbitrageur specializes in location but not on vehicles. He or she buys different types of vehicles from the exact same locations.

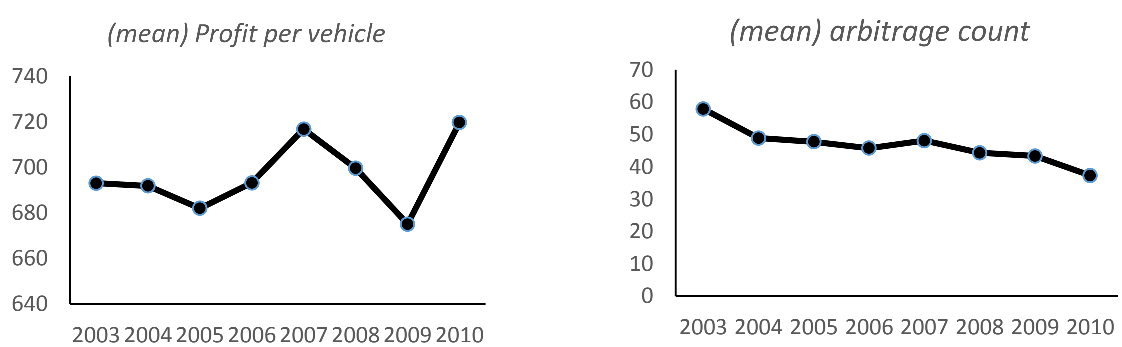
Figure 3.1. This figure depicts two different arbitrageur strategies.

[24] J D Power & associates is a world renowned marketing firm that conducts surveys on car and vehicle qualities. J D Power & associates categorize vehicles based on body types as SUV, Sedan, Truck, Convertible, Coupe, Wagon, Van, Crossover or Hatchback. Another categorization of vehicles in our sample is to use vehicle makes which include Ford, Acura, Fiat, Kia, Nissan, Suzuki, Toyota, Chevrolet, etc.. Our key results are robust to both types.

3.4.2.2 Arbitrage Profits

Arbitrage profits are calculated, as described in Table 3.1 below.

Figure 3.2 below shows how mean arbitrage profits and arbitrage count vary over time. The mean arbitrage profits are adjusted for inflation using the consumer price index for the corresponding year. We observe that while the mean arbitrage profits fluctuate, the arbitrage count declines over time. The first dependent variable (the mean arbitrage profit of the arbitrageur per year) can be used to determine the effectiveness of the arbitrageur in the market. This corresponds to Fig 3.2 (a). However, the mean profits per vehicle are in themselves insufficient to study arbitrage behavior. Consider an arbitrageur who makes \$2,000 in arbitrage profits per vehicle. This arbitrageur is more effective than one who makes \$2,000 profit by arbitraging 10 vehicles. Therefore, it is important to study antecedents that influence the number of vehicles arbitraged, or arbitrage intensity in the market. The number of vehicles arbitraged also indicates whether the arbitrageur has been able to exploit inefficiencies in the market, as well as whether the arbitrageur has been able to adapt his or her behavior to arrive at the right strategy for the market. The change of arbitrage intensity over time is depicted in Fig 3.2 (b) below.



a) mean *ArbitrageProfit* by year

b) mean *NumArbitraged* by year

Figure 3.2: Graphs depicting of arbitrage profits(mean*) and arbitrage count(*mean*) by year*

As noted above, we identify 720 arbitrageurs who have arbitraged at least 50 vehicles over eightyears between 2003 and 2010. We constructed a two-dimensional panel of observations for each arbitrageur *i* during year *t*. We tested the effect of vehicle specialization and source location specialization on mean arbitrage profits earned by the arbitrageur. We note that $VehicleSpecialization_{it}$ and $LocationSpecialization_{it}$ affect the mean arbitrage profit per vehicle and are the focal independent variables. The limits of arbitrage theory suggest that capital limitations affect arbitrage activity. Hence, we include the mean capital available to the arbitrageur ($CapitalWk_{it}$) as a proxy to measure the capital that is available to the arbitrageur in a given period. Note that this measure is neither accurate nor complete, since we do not know the amount of capital accessible to the arbitrageur at any point in time. Given the fact that in our definition of arbitrage, we use 7 days to indicate a flip; we assume that the mean rollover time for each vehicle being arbitraged (or sold) from a dealer's lot is 7 days. This means that the average capital available to the arbitrageur is equivalent to the total capital spent on purchasing vehicles from the market in a particular week[25]. We calculate the total number of arbitrage opportunities available within a radius of 500 miles from the arbitrageur ($totalOpportunities500_{it}$). Chapter 2, section 2.5.2 discusses how the number of opportunities is derived. This variable is used as an indicator of opportunities in the market of which the arbitrageur should be aware. These data are not available to the arbitrageur as he or she does not have visibility into all trades happening at a given point in time. The arbitrageur also cannot calculate the profits ex-ante, if such transactions were to happen. We assume that

[25] The total capital available to a dealer can be thought of as the total Cash in Hand + Credit accessible + Value realized from Used Car Inventory Turnover at any point in time. It is difficult to estimate cash in Hand or Credit accessible to the dealer. These depend on a variety of factors including size of the dealership, size of the business etc. However, we can use CapitalWk as a proxy for Used Car inventory turnover. Used Car Inventory turnover varies by dealer, location and time of year for sale. The general consensus amongst all dealers is that vehicles have to sell within 30 days (http://www.vauto.com/dealer-resources/managing-used-age/avoid-used-vehicle-inventory-age-problems/). However, this cannot be assumed true. Inventory turnover would potentially be difficult to calculate and this measure can be improved in further versions of the paper. We currently use CapitalMonth and CapitalYear to indicate the total capital spent by the arbitrageur in the market to purchase vehicles which are also proxies. The MaxCapitalWk, indicating maximum capital available to the arbitrageur, is also calculated but is subject to the outlier effect.

arbitrageurs who are aware of the market's operation will use this information to increase their trade, and, hence, we include this variable in the model. We include (*VehicleMileage*$_{it}$ or *VehicleValuation*$_{it}$) as these variables control for the value of the car. The higher the value of the vehicle, the higher the risk to the arbitrageur and possibly higher the profit possibility per vehicle It is also possible that more expensive vehicles bring in less profit due to limited demand. Electronic commerce is a major component of the market, providing significantly increased reach for an arbitrageur. In some ways sourcing via e-commerce enables arbitrageurs to shift their demand to locations where they can find a matching vehicle in the expected price range. The percentage of vehicles sourced from the e-commerce channel (pctArbitragedeChannel$_{it}$) indicates the extent to which mean arbitrage profits are affected by sourcing from e-commerce channels. The proportion of vehicles (*ProportionArbitraged*$_{it}$*i*) arbitraged by an arbitrageur indicates the extent to which arbitrage is considered an important constituent of the arbitrageur's overall strategy for arbitrage. γ_t is used to indicate the time fixed effects. α is the intercept and U_i is used to indicate user level fixed effects.

Our two-dimensional panel model to test H1 is specified below and descriptive statistics of key variables are discussed in Table 3.1 below.

$$\text{MeanArbitrageProfit}_{it} = \alpha + U_i + \beta_1\text{VehicleSpecialization}_{it} + \beta_2\text{LocationSpecialization}_{it} + \beta_3\text{CapitalWk}_{it} + \beta_4\text{NumVehiclesPurchased}_{it} + \beta_5\,\text{MeanVehicleMileage}_{it} + \beta_6\,\text{totalOpportunities500}_{it} + \beta_7\,\text{pcteAccess500}_{it} + \beta_8\text{proprotionArbitrage}_{it} + \beta_9\text{pctArbitragedeChannel}_{it} + \gamma_t + \epsilon_{it}$$

Table 3.1: Descriptive statistics of key variables

Variable Name	Description	Descriptive statistics	
		Mean (stdev)	Min,max
meanArbitrageProfit$_{it}^{a}$	Mean arbitrage profit per vehicle of arbitrageur *i* in year *t* divided by 10000.(This profit is adjusted for inflation using the consumer price index http://www.bls.gov/cpi/). Profits for each transaction are calculated as follows: ArbitrageProfit = salePriceatDestination - salePriceatSource – BuyerFees – SellerFees – transportationFees. Transportation fees are calculated as described in Appendix A.	0.069 (0.04)	-0.234, .699
VehicleSpecialization$_{it}^{b}$	Mean Gini coefficient indicating specialization of the arbitrageur *i* with respect to vehicle power type in year *t* for the arbitrgeud vehicles	0.329 (0.199)	0, 0.80
LocationSpecialization$_{it}$	Mean Gini coefficient indicating specialization of the arbitrageur *i* with respect to source auction locations in year *t* for the arbitraged vehicles.	0.265 (0.208)	0, 0.83
NumVehiclesPurchased$_{it}$	Count of vehicles purchased by arbitrageur *i* divided by 100 in year *t*.	362.40(442.24)	2, 6071
NumArbitraged$_{it}$	Count of vehicles arbitraged by the arbitrageur *i* in year *t*.	46.92(113.1)	1, 2771
proportionArbitrage$_{it}$	Proportion of vehicles arbitraged by arbitrageur *i* in year *t*.	0.136 (0.12)	.0003,1
pctArbitragedeChannel$_{it}$	Proportion of vehicles that were arbitraged, which were sourced on the electronic channel (i.e. either webcast or standalone electronic market) by arbitrageur *i* in year *t*.	0.040 (0.148)	0,1
meanVehicleMileage$_{it}$	Mean mileage of all vehicles purchased by arbitrageur *i* divided by 10000 in year *t*.	6.52 (3.01)	1.27, 7.71
*totalVehiclesAccess500*it^{c}	Count of vehicles accessible to arbitrageur *i* within a radius of 500 miles divided by 10000 in year *t*	107.7 (48.13)	7.90, 233.3
totalOpportunities500$_{it}^{d}$	Count of opportunities for arbitrage within 500 miles of arbitrageur *i* in year *t* divided by 10000	58.09 (37.96)	1.70, 204.90
pcteAccess500$_{it}^{e}$	Count vehicles accessible on the electronic channels to arbitrageur *i* within a radius of 500 miles divided by 10000 in year *t*.	0.819 (0.22)	0.177,1
CapitalWk$_{it}^{f}$	Capital available to the arbitrageur *i* during year *t* divided by 10000. (This is calculated as the mean capital spent spent during a particular week by the arbitrageur *i* in a particular year *t*).	7.42 (9.79)	0.130, 165.92

a,b – This measure is tested for Gini (as well), and is tested with Power type as well
C,d,e – similar measures are created for a radius of 250 miles, 200 miles. Results are found to be robust
f- This measure is created for capital available in a year, capital available in a month, and the maximum capital available for a particular week. Results are found to be similar and robust.

Table 3.2 Effect of Specialization on meanArbitrageProfit

	meanArbitrageProfit
VehicleSpecialization	0.0166***(0.0044)
LocationSpecialization	0.0085**(0.0034)
CapitalWk	-0.0000(0.0002)
NumVehiclesPurchased	0.0000(0.0000)
meanVehicleMileage	-0.0039***(0.0006)
totalOpportunities500	0.0001*(0.0000)
pcteAccess500	0.0113(0.0230)
proportionArbitrage	-0.0003(0.0083)
pctArbitragedeChannel	0.0168***(0.0045)
Time Fixed Effects	*Included*
User level Fixed Effects	*Included*
Constant	0.0741***(0.0092)
N	4000
R^2	0.601
adj. R^2	0.511
F-stat	8.2679
Degree Free. Min	16.0000
Degree Free. R	3264.0000

Standard errors in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)
(*MeanArbitrageProfit is adjusted for inflation using the consumer price index, as noted above*)

From the above results, we see that the vehicle specialization and auction specialization play important roles in determining both intensity and volume of arbitrage. We further estimate that a single standard deviation increase in vehicle specialization increases the mean arbitrage profit earned by the arbitrageur by 4.13 % (or $28.7). Similarly, a single standard deviation increase in auction specialization increases mean arbitrage profits by 2.44% (or $17). Similarly, if the capital accessible to the arbitrageur is increased by one standard deviation (i.e., $97900), we find that the mean arbitrage profits are not affected. Also, a single standard deviation increase in number of vehicles sourced using the electronic channels would increase the mean profits by 3.5% (or ~$25). This means that the arbitrageur is more likely to source cheaper vehicles from remote locations using the electronic channels, which would increase the profits. Another interesting effect is that of mean vehicle mileage. An increase of one standard deviation of vehicle mileage reduces mean profits by 16% (or $117). An increase of one standard deviation in the total number of cars purchased increases the number of vehicles arbitraged by 1,122% (or 57 vehicles).

Our results indicate that H1 is supported. Specialization affects arbitrage profitability. We test for robustness of the next section.

3.4.2.2 Tests for Robustness

For robustness, we use the random effects model to test whether the results vary given random effects for the arbitrageur (Refer Table C1 in Appendix C for results). The Hausman tests for model 1 finds no systematic differences (χ^2 (15) = 31.09, p > 0.01) between fixed effects (which are always consistent but efficient) and random effects estimates (which are consistent but may be inefficient), which proves that the results do not change. In addition, to test for the variations in measurements, we alter the measurement of specialization to the Theil Index and show that these results hold true even with the Theil index (Ref Table C3 in Appendix C for results). We use vehicle makes instead of the J.D. Power and Associates vehicle category for vehicle specialization. The results are consistent with the vehicle makes as well (Ref. Table C2 in Appendix C for results).

3.4.3 Testing H2: Arbitrageur Specializations evolve over a period of time

We plot the set of 720 arbitrageurs based on their overall specializations for the time period of our study (2003 – 2010). Figure 3.3 below depicts these 720 arbitrageurs split into 3 groups, or terciles, based on the number of vehicles arbitraged. We see that the greater the number of vehicles arbitraged (or the higher the intensity of arbitrage), the higher the specialization will be on both dimensions. For high volume arbitrageurs, the vehicle specialization is higher than that of the low volume arbitrageurs. Similarly, for the higher volume arbitrageurs, location specialization is higher than location specialization for low-volume arbitrageurs. In the next section, we study how arbitrageur strategies evolve.

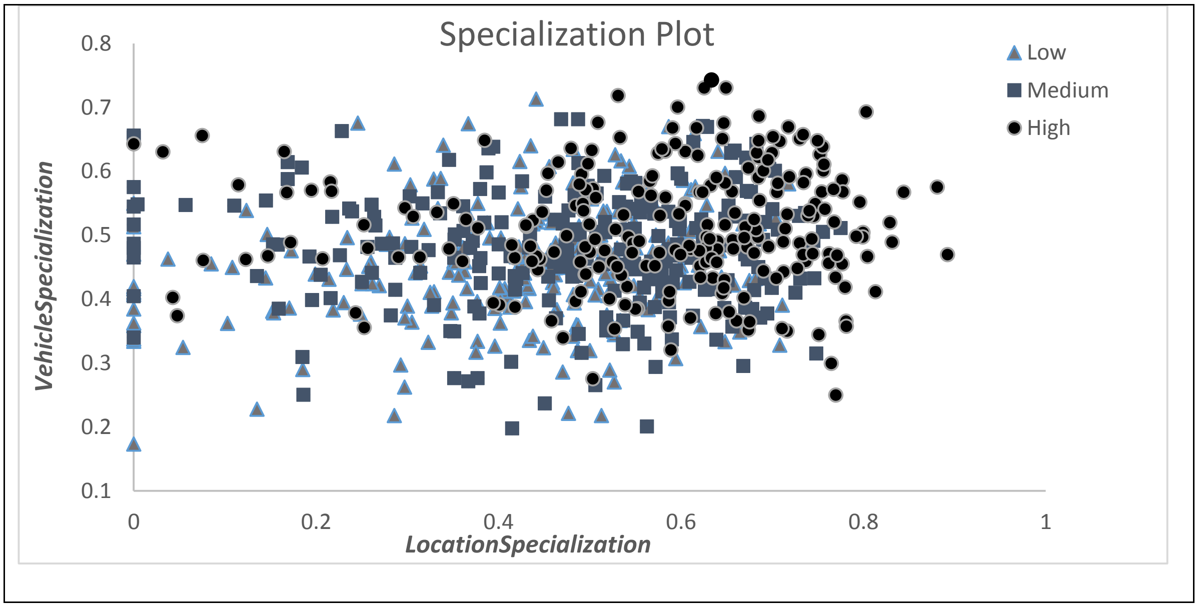


Figure 3.3: Arbitrageurs are divided into terciles (3 groups) based on the count of vehicles arbitraged between 2003 and 2010. The specializations are axis of VehicleSpecialization (Y-axis) vs LocationSpecialization (X-axis).

Table 3.3 Describing the distribution of terciles amongst the quadrants above

Total = 720	*Count*	*Count*	*Count*	*Count*
Count based category VehicleSpecialization-LocationSpecialization	Q1 (LOW-LOW)	Q2 (LOW-HIGH)	Q3 (HIGH-LOW)	Q4 (HIGH-HIGH)
Tercile 1 (bottom 33rd ercentile)	108	43	48	42
Tercile 2 (33rd to 66th percentile)	52	59	69	59
Tercile 3 (Top 33rd percentile)	25	30	71	114

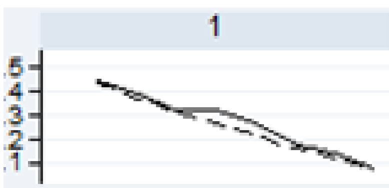
We see that a considerable number of arbitrageurs specialize in both dimensions in tercile 3, indicating that the high volume arbitrageurs are mostly specialists, or that arbitrageurs who engage in a large number of arbitrages specialize more. In the next section we examine how arbitrageurs evolve, and we categorize arbitrageurs into different groups based on their specialization.

3.4.3.1 Arbitrageur Evolution

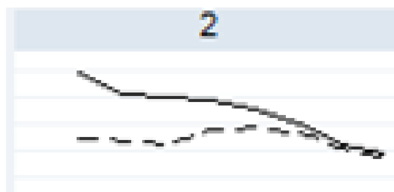
In order to examine how arbitrageur specialization strategy evolves over time, we first classify arbitrageurs based on how *LocationSpecialization* and *VehicleSpecialization* change over time. For each arbitrageur i during year t, we construct a Gini coefficient that is indicative of vehicle specialization. This indicates an arbitrageur's preference for vehicles belonging to a particular J.D. Power and Associates category ($VehicleSpecialization_{it}$). Similarly, for each period t, we construct a Gini coefficient that indicates the distribution of source locations from which arbitrageur i sourced vehicles ($LocationSpecialization_{it}$)[26]. Our panel consists of 4,000 observations with these two specialization variables (i.e., $VehicleSpecialization_{it}$ and $LocationSpecialization_{it}$), where t indicates the time periods and i indicates unique vehicle identifiers. Off these 720 arbitrageurs, only 96 arbitrageurs engage in arbitrage for all eightyears of our study. Thus, because of the missing values, using a clustering algorithm such as k-means or Ward's hierarchical clustering algorithm to determine groups of arbitrageurs based on the specialization variables would reduce the clustered dataset[27] significantly. In order to overcome this limitation, we use the slopes of specialization indicators (i.e. $VehicleSpecialization_{it}$ and $LocationSpecialization_{it}$) to construct categories of arbitrageurs. We use the models $VehicleSpecialization_{it} = \alpha + \beta_{0i}.\ (t - 2000)$ and $LocationSpecialization_{it} = \alpha + \beta_{1i}.(t - 2000)$ to obtain the slope coefficients (i.e. β_{0i} *and* β_{1i}) for each arbitrageur i. This will enable us to classify arbitrageurs based on the directionality of change in specialization over time. If β_{0i} is positive (negative), then the arbitrageur i became more (less) specialized with respect to vehicles. Similarly, if β_{1i} is positive (negative), then arbitrageur i became more (less) specialized over time with respect to sourcing locations. We divide the arbitrageur population into three groups (terciles): *G1,G2* and *G3* based on β_{0i} i.e. slope of *VehicleSpecialization.* Next, we divide the arbitrageur population into three groups (terciles): *L1*, *L2* and

[26] Vehicle specialization is indicated mainly by the specialization based on the power of the vehicle. Similarly, to test for robustness we also use the Vehicle Make e.g. Ford to indicate specialization.
[27]Refer Appendix D

L3 based on β_{1i} i.e. slope of *LocationSpecialization*. We then construct 9 unique groups based on two different tercile sets: *G1, G2, G3* and *L1, L2, L3*. Each arbitrageur belonging to a group belongs to one of the groups *G1, G2* and *G3* and to one of the groups *L1, L2* and *L3*. These 9 groups uniquely identify different evolution patterns exhibited by arbitrageurs with respect to their specialization(s). The mean values of $VehicleSpecialization_{it}$ and $LocationSpecialization_{it}$ by year are plotted in Figure 3.4 below for each of the 9 groups. A discussion using the k-means clustering algorithm is presented in Appendix D. We also discuss the usage of better-fit equations of the second orde to derive these groups using clustering algorithms.

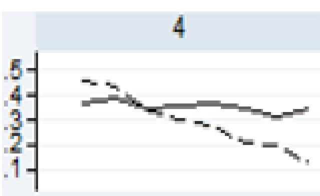


Group 1: Arbitrageurs become less specialized with respect to both vehicles and source locations. N = 148

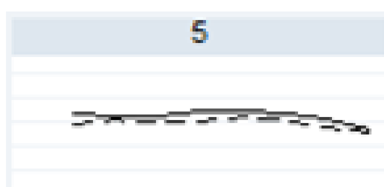


Group 2: Arbitrageurs become less specialized with respect to vehicles but remain low specialists with respect to source locations. N = 53

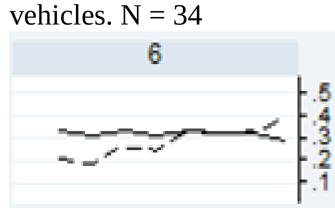
Group 3:Arbitrageurs belonging to this group become more specialized with respect to source locations and become less specialized with respect to vehicles. N = 34



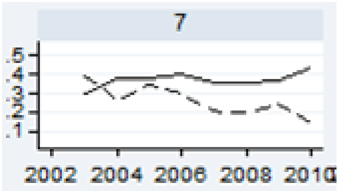
Group 4: Arbitrageurs become less specialized with respect to source locations, while remaining high vehicle specialists. N= 53



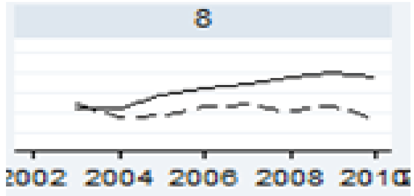
Group 5: arbitrageurs do not change their specializations. N=123



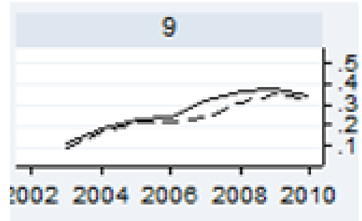
Group 6: Arbitrageurs increase source location specialization but remain low vehicle specialists N=58



Group 7: These arbitrageurs remain constantly high on vehicle specialization, but become increasingly less specialized on auction specialization. N = 34



Group 8: These arbitrageurs become high vehicle specialists, but remain low location specialists.
N=58



Group 9: These arbitrageurs become increasingly specialized on both the vehicle and source location.
N= 142

Figure 3.4: Groups of Arbitrageurs describing the evolution of arbitrageur specialization in time. The dark line indicates mean VehicleSpecialization in year t. The dotted line indicates LocationSpecialization in year t.

In order to test for robustness in these patterns, we keep only 394 arbitrageurs who have arbitraged vehicles in at least 6 years out of eight(between 2003 and 2010). Thus, H2 is supported. Arbitrageur specialization behavior changes over time and in different ways. This result is important in the light of the prior literature on limits of arbitrage and efficient markets hypothesis. Firstly, the efficient markets hypothesis states that when markets become inefficient, multiple traders quickly consume these inefficiencies. The behavioral finance scholars differ in that they claim that only a few specialized traders would engage in arbitrage, and they are limited. We see from our results above that only 720 traders out of 12,196 actually engage in arbitrage, which confirms the predictions of behavioral finance. These groups confirm that are also specialists in the sense that their specialization affects profits and arbitrage intensity. Secondly, these 720 arbitrageurs show different behaviors as predicted in the Adaptive Markets Hypothesis since several arbitrageurs exhibit common patterns of evolution.

From Figure 3.4 above, we see that arbitrageur strategies are neither constant nor uniform, but that they evolve in very specific ways in markets. In the next section we address the following question: *What factors affect specialization?* We test for antecedents that alter an arbitrageur’s strategy over time.

From the above groupings of arbitrageurs based on the directionality of the evolution of their specializations, we see that arbitrageur groups can either increase or decrease their specializations in both dimensions.

3.4.4 Testing H3, H4 and H5: Antecedents influencing arbitrageur strategy

From H2, we recognize that arbitrageurs specialize over time in both the vehicle type and sourcing location.[28]. There are two problems when we analyze this: Both *VehicleSpecialization* and *LocationSpecialization* are endogenous to *NumArbitraged*, because factors that affect the intensity of arbitrage also affect the change of specialization, and vice-versa. Both *Specialization* and *NumArbitraged* could be changing together introducing simultaneity bias. Since *MeanArbitrageProfit* is influenced by *NumArbitraged* and the specialization variables, we model arbitrage outcomes as a system of equations shown below, and instrument *LocationSpecialization and VehicleSpecialization*. The 4 models that describe our system are shown below. We further use a 3-Stage least squares estimator to obtain the effects of specialization on arbitrage outcomes. This accounts for the simultaneity bias and endogenity with respect to specialization and the outcomes of arbitrage.

i. $MeanArbitrageProfits_{it}=f(ProportionArbitrage_{it},VehicleSpecialization_{it},LocationSpecialization,Profitcontrols_{it})$

ii. $NumArbitraged_{it} = f\,(VehicleSpecialization_{it}, LocationSpecialization_{it}, Countcontrols_{it})$

The specialization variables *LocationSpecialization* and *VehicleSpecialization* are defined as follows:

iii. $VehicleSpecialization_{it}= f\,(ProportionArbitrage_{it,}, OverallVehicleSpecialization_{it}, Vehiclecontrols_{it})$

iv. $LocationSpecialization_{it}= f\,(ProportionArbitrage_{it,},OverallLocationSpecialization_{it,,}Locationcontrols_{it})$

The econometric specifications are four fixed effects models as follows:

$$\text{MeanArbitrageProfit}_{it} = \alpha + U_i + \beta_1\text{VehicleSpecialization}_{it} + \beta_2\text{LocationSpecialization}_{it} + \beta_3\text{CapitalWk}_{it} + \beta_4\text{NumVehiclesPurchased}_{it} + \beta_5\,\text{MeanVehicleMileage}_{it} + \beta_6\,\text{totalOpportunities500}_{it} + \beta_7\,\text{pcteAccess500}_{it} + \beta_8\text{proprotionArbitrage}_{it} + \beta_9\text{pctArbitragedeChannel}_{it} + \gamma_t + \epsilon_{it} \quad \text{--}(1)$$

[28] An Alternative method to test the evolution of specialization using state-based transitions is described in Appendix F below. This method though robust discretizes the continuous variables for specialization. It is also difficult to interpret since simulations need not be exact.

$NumArbitraged_{it} = \alpha + U_i + \beta_1 VehicleSpecialization_{it} + \beta_2 LocationSpecialization_{it} + \beta_3 CapitalWk_{it} + \beta_4 NumVehiclesPurchased_{it} + \beta_5\ MeanVehicleMileage_{it} + \beta_6\ totalOpportunities500_{it} + \beta_7\ pcteAccess500_{it} + \gamma_t + \epsilon_{it}$ --(*2*)

$VehicleSpecialization_{it} = \alpha + U_i + \beta 1 OverallVehicleSpecialization_{it} + \beta_2\ proprotionArbitrage_{it} + \beta_3 CapitalWk_{it} + \beta_4 NumVehiclesPurchased_{it} + \beta_5\ MeanVehicleValuation_{it} + \beta_6 totalVehiclesAccess500_{it} + \beta_7\ pcteAccess500_{it} + \beta_8 pctArbitragedeChannel_{it} + \gamma_t + \epsilon_{it}$ -- (*3*)

$LocationSpecialization_{it} = \alpha + U_i + \beta_1 OverallLocationSpecialization_{it} + \beta_2 proprotionArbitrage_{it} + \beta_3 CapitalWk_{it} + \beta_4\ totalVehiclesAccess500_{it} + \beta_5\ pcteAccess500_{it} + \beta_6\ pctArbitragedeChannel_{it} + \beta_7\ cntAuction500_i + \gamma_t + \epsilon_{it}$ ---(*4*)

Based on the above specifications, we have two systems of equations based on the focal dependent variables indicating arbitrage outcomes (*MeanArbitrageProfit* and *NumArbitraged*). They are (1,3,4) and (2,3,4). We apply 3SLS for these sets of equations.

Model (1) is the same as the one described in section 3.4.2. Model (2) depicts variables correlated with the number of arbitraged vehicles. Model (3) depicts independent variables correlated with *VehicleSpecialization* and Model (4) depicts variables correlated with *LocationSpecialization*. We describe our instrument variable below.

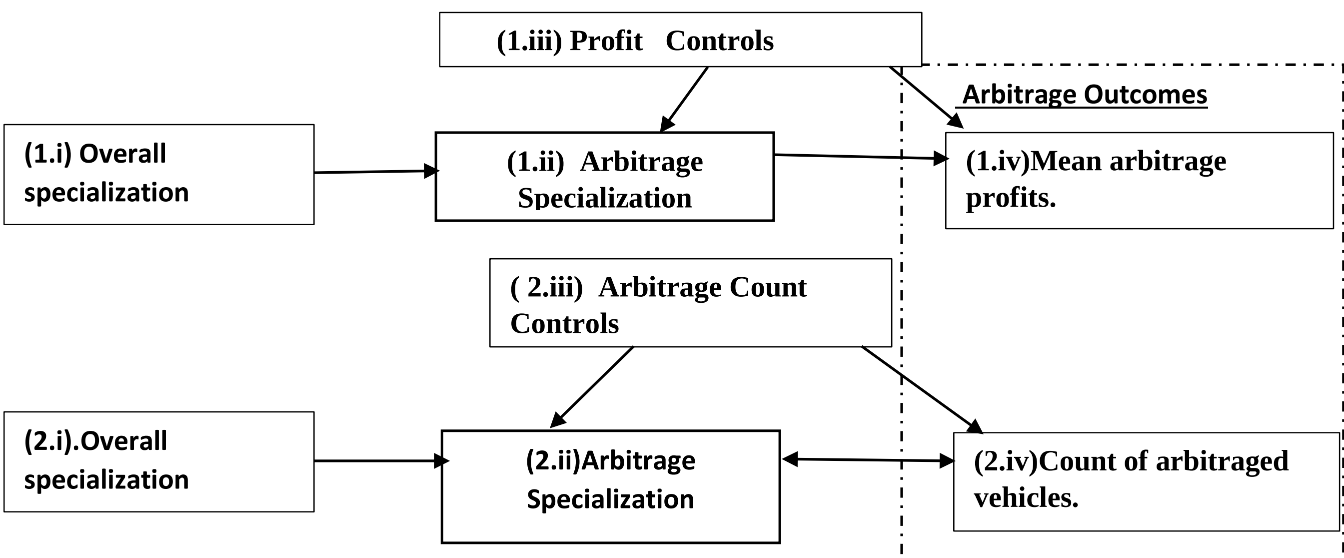


Figure 3.5 Diagram indicating system dependence between Specialization, Controls and Arbitrage outcomes. The path 1.x indicates the system of equations 1,3,4 and the path 2.x indicates the system of equations 2,3,4

3.4.4.1 Instrument: As mentioned in section 3.2.4, every arbitrageur is also a dealer. Arbitrageurs also purchase vehicles to their dealer lot, which are not arbitraged. From table 3.1 we see that the mean of *proportionArbitrage* is 0.13 and the standard deviation is 0.12, which shows that most arbitrageurs arbitrage less than 15% of their total inventory. (There is also only 1 arbitrageur amongst 720 who has arbitraged 100% of his inventory of 5 vehicles in a particular year.) Thus, the choice of vehicles for arbitrage might depend on the arbitrageur's overall choice preferences. We believe that the overall behavior (specialization) of the arbitrageur, calculated as the Gini coefficient of all vehicles purchased by an arbitrageur in a particular year (*OverallVehicleSpecialization*), and the Gini coefficient of all source locations visited by the arbitrageur in a particular year, indicated by (*OverallSourceLocationSpecialization*), affects the arbitrageur's specialization variables with respect to vehicles arbitraged. The *OverallVehicleSpecialization* variable affects the *vehicleSpecialization* for those vehicles that are arbitraged, and the specialization with respect to vehicles that are not arbitraged. Similarly, the *OverallLocationSpecialization* variable affects the location specialization for those vehicles that are arbitraged, and those that are not arbitraged. These two variables can be thought to be indicators of the overall location strategy of the arbitrageur influencing the arbitrageur's *LocationSpecialization* with respect to arbitrage. We note that the arbitrage behavior of an arbitrageur depends on the overall behavior of the arbitrageur and his or her strategy is significantly affected by biases he or she exhibits while purchasing vehicles for his lot. For example, a Ford dealer would most likely purchase Ford SUVs as opposed to Toyota SUVs or GM's SUVs for the dealership. (He or she is also most likely to arbitrage Ford SUVs since s/he knows the market very well.) Thus, we claim that an arbitrageur's overall strategy influences arbitrage strategy, which in turn influences the profits earned in the market. Similarly, the overall behavior (of specialization) does not affect the arbitrageur's mean profits or the number of vehicles arbitraged, which are functions of only those vehicles arbitraged. We use specialization indicators for vehicles and source locations for all vehicles purchased by an arbitrageur in a given year as the key instrumental variables.

OverallVehicleSpecializationl and *OverallLocationSpecialization* affect *VehicleSpecialization* and *LocationSpecialization*, but do not affect the *MeanArbitragedProfit* or *NumArbitraged*, which are dependent only on the vehicles sold. This is because both *meanArbitrageProfit* and *NumArbitraged* are affected only by the sale of vehicles purchased for arbitrage (i.e., *VehicleSpecializaiton*). The summary statistics of the key variables are given below in Table 3.4. The rest of the variables are described in Table 3.1.

Table 3.4. Descriptive Statistics of key instrument variables

Variables	Description	mean(SD)	min (max)
$OverallVehicleSpecialization_{it}$	mean Gini coefficient indicating specialization of the arbitrageur *i* with respect to vehicle power type in year *t* for all vehicles purchased in the year.	0.25(0.254)	0(0.81)
$OverallLocationSpecialization_{it}$	mean Gini coefficient indicating specialization of arbitrageur *i* with respect to source locations in year *t* for all arbitraged vehicles.	0.45 (0.191)	0(0.87)

Table 3.5 Results of the 3SLS regression

Models	(1,2,4)	(2,3,4)
Key Dependent Variables(*Model i,ii*)	*meanArbitrageProfit*	*NumArbitraged*
VehicleSpecialization	0.0437*(0.0260)	182.9287***(27.6703)
AuctionSpecialization	0.0390**(0.0197)	18.6377(30.0608)
CapitalWk	-0.0002(0.0002)	4.6164***(0.3433)
NumCarsPurchased	-0.0000(0.0000)	0.1214***(0.0072)
proportionArbitrage	-0.0527***(0.0196)	-
pctArbitragedeChannel	0.0164***(0.0054)	-
meanVehicleMileage	-37.6251***(5.6419)	4736.9521(8632.8368)
totalOppurtunities500	0.0001*(0.0000)	-0.2130***(0.0661)
pcteAccess500	0.0102(0.0211)	-6.6477(39.2015)
c.CapitalWk#c.pctArbitragedeChannel	-0.0001(0.0003)	-3.8364***(0.3976)
Time specific Coefficients	Included	Included
Arbitrageur specific Coefficients	Included	Included
Intercept	0.0579***(0.0173)	-69.6791**(30.8655)
VehicleSpecialization (*Model iii*)		
proportionArbitrage	1.0015***(0.0278)	1.0522***(0.0271)
OverallVehicleSpecialization	0.3270***(0.0305)	0.1959***(0.0250)
pcteVehicles	0.0868(0.0629)	-
CapitalWk	0.0050***(0.0004)	0.0053***(0.0004)
totalVehiclesAccess500	-0.0000(0.0006)	0.0003(0.0005)
pcteAccess500	-0.0266(0.0965)	-0.0206(0.0926)
meanVehicleValuation	0.0000(0.0000)	-0.0000(0.0000)
Intercept for VehicleSpecialization	0.0369(0.0791)	0.0867(0.0762)
Time specific Coefficients	Included	Included
Arbitrageur specific Coefficients	Included	Included
LocationSpecialization (*Model iv*)		
proportionArbitrage	0.8065***(0.0353)	0.8112***(0.0353)
OverallLocationSpecialization	0.2091***(0.0176)	0.2169***(0.0175)
pcteVehicles	0.0129(0.0798)	0.0278(0.0806)
CapitalWk	0.0051***(0.0006)	0.0051***(0.0006)
totalVehiclesAccess500	0.0006(0.0007)	0.0005(0.0007)
pcteAccess500	-0.0046(0.1227)	-0.0171(0.1230)
cntAuction500	-0.0167**(0.0072)	-0.0172***(0.0065)
Intercept for LocationSpecialization	0.4235***(0.1320)	0.4327***(0.1258)
Time specific Coefficients	Included	Included
Arbitrageur specific Coefficients	Included	Included
N	3977	3977
R^2	0.578	0.816
Degree Free. min	725.0000	723.0000
log-likelihood.	1.40e+04	-1.50e+04
Chi-squared	5800.5377	1.75e+04

Standard errors in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

3.4.4.2 Interpretation of results

H3 is supported. For both *VehicleSpecialization* and *LocationSpecialization*, the *proportionArbitraged* variable is positive and statistically significant at the 0.001 levels. Similarly for *MeanArbitrageProfit* and the *proportionArbitraged* is positive and significant. What this shows is that intensity of arbitrage affects both the Specialization variables. If the *proportionArbitraged* increases by 1 standard deviation, then VehicleSpecialization increases by 36% - 39.5% or between 0.12 – 0.14. Similarly, *LocationSpecialization* increases between 63% -66% or 0.16-0.18 Thus, as arbitrageurs increase their involvement in markets, by arbitraging more of their inventory in markets, they seemingly become more specialized. This finding is in sync with the theory that specialization helps arbitrageurs sustain profitability in markets and the general belief that arbitrageur specialization increases as arbitrageurs continue to engage in markets.

H4 is NOT supported. We see that the number of vehicles sourced from the electronic channels (or the number of vehicles available on the electronic channels within a radius of 500 miles) do not affect specialization in a significant way. However, the percentage of vehicles sourced from an electronic channel and later arbitraged affects arbitrage profits as explained above. This is because electronic trading channels provide increased transaction immediacy enabling arbitrageurs to locate vehicles priced optimally Ref. Chapter 2). H5 is supported. As the amount of capital availability increases for arbitrage, arbitrageurs can fundamentally increase their involvement in markets without worrying about profits. However, if their main goal is to make profits, they tend to become more specialized as they increase their purchase in markets. A single standard deviation increase in capital available for arbitrage increases *vehicleSpecialization* by 15%, or 0.04. Similarly, a one standard deviation increase in capital available for arbitrage increases the *locationSpecialization* by 18%, or 0.05. Thus, capital availability influences specializations positively. Capital also the intensity of arbitrage by increasing an arbitrageur’s participation in markets. This shows that capital availability is one of the most influential factors affecting arbitrage (both with respect to arbitrageur strategy and with respect to arbitrageur availability).

Testing the Over-identification condition: We run the Sargan’s test (Lung-Fei 1992) to test for over-identification on both simultaneous equation models (1,3,4) and (2,3,4). We find that for the model

(1,3,4) the Hansen-Sargan over-identification statistic is 249.562 and the $\chi^2(10) = 0.000$. This shows that the instruments are valid. However, for (2,3,4) the Hansen-Sargan over-identification statistic is 281.669 and the $\chi^2(22) = 0.000$. This shows that the instruments are valid.

3.5 Conclusion

We study behaviors of arbitrageurs and test various predictions of the limits of arbitrage (see Shleifer and Vishny 1997) and the adaptive markets hypothesis (see Lo 2004) in the context of spatial arbitrage. Our analysis supports the assumption that specialization matters for arbitrageurs and influences the extent to which arbitrageurs make profits. We also show that specialization determines the intensity of arbitrage.

To our knowledge, we are the first to study arbitrageur strategy with respect to sourcing specializations on two dimensions—asset specialization and source location specialization—by observing and identifying specific arbitrageur behaviors. We find that arbitrageur specialization on two dimensions—namely, asset type and sourcing locations matters in this market.

Firstly, we show that arbitrageur specialization affects arbitrage profits and intensity (i.e., number of vehicles arbitraged). Secondly, we show that many different types of specialization strategies exist in this market. We identify nine predominant strategies and the corresponding arbitrageurs who follow those strategies. This finding is in sync with the adaptive markets hypothesis, which predicts the presence of different groups of arbitrageurs, each with different behaviors in markets. We then analyze the antecedents that affect arbitrage strategy. We show that as arbitrageurs increase arbitrage activity in markets, they tend to become more specialized on both the source location and vehicle types. While vehicles sourced on the electronic commerce channels influence profits positively, we are unable to find any effect of electronic trading (or access on electronic vehicles) on specialization. This is in sync with our earlier assumptions and conclusions in the limits of arbitrageur literature; our results show that capital available for arbitrage increases specialization and also directly influences the intensity of arbitrage.

Future work on this paper includes improving measures on capital availability by taking into account factors such as the 2008 depression and macro-economic factors in markets.

3.5.1 Limitations and future research

While our predictions and analysis test the limits of arbitrage theory and, to a certain extent, support predictions of the Adaptive Markets Hypothesis, our scope is limited to “vehicles” and spatial arbitrage of “vehicles.” That being said, the automobile market and the used automobile market in the United States forms a considerable proportion of the US GDP (approximately 4%) and is large enough to be studied on its own. Another limitation could be the measures of specialization. We chose the Gini coefficient (and test for robustness with the Thiel Index). However, almost all measures are extremely sensitive to the count of items in a distribution. Antecedents could change under conditions when the goods traded are commodities other than vehicles. Future research opportunities could include testing our hypothesis with different good types or commodities such as food grains or heavy metals that are traded in physical markets on a wholesale basis.

Chapter 4: CONCLUSION

Spatial arbitrage and arbitrageurs are fundamental to markets since they enable matching supply and demand across locations, and, therefore by enabling convergence of prices across geography make markets efficient.(Barrett 2008, Takayama and Judge 1964). The two essays in this dissertation enhance our understanding about spatial arbitrage and arbitrageur sourcing behaviors.

In the first essay, we make three key contributions. Firstly, we argue that spatial arbitrage is a measure of market efficiency and has several advantages over price dispersion. Secondly, we empirically examine the effect of two different modes of electronic commerce channels on spatial arbitrage. We show that channel features—namely, reach and transaction immediacy—affect the extent to which arbitrageurs are able to use the corresponding channel to exploit arbitrage opportunities. Finally, we document factors that determine the arbitrageur's choice of source locations. Our main key finding is that the effect of electronic commerce is nuanced: As markets get more efficient and spatial arbitrage (and arbitrage opportunities) is reduced, arbitrageurs also get better at exploiting inefficiencies that remain.

We identify that the webcast channel that has high reach but no transaction immediacy results in a reduction of spatial arbitrage opportunities for the arbitrageur. However, the standalone electronic market that has high reach and high transaction immediacy increases the arbitrageur's ability to identify and exploit such opportunities. Further, we analyze arbitrageur sourcing behaviors in markets and find various interesting facts about how arbitrageurs choose sourcing locations. We hope that the arguments regarding spatial arbitrage and market efficiency are convincing enough to make spatial arbitrage a powerful alternative measure of market efficiency. We believe that in most markets where identifiers of goods, traders, and locations are available, spatial arbitrage could be identified easily. We hope that future scholars in information systems, finance, and economics will use spatial arbitrage as a valid measure of market efficiency in their studies.

In the second essay, we confirm theoretical predictions pertaining to factors stated in the limits of arbitrage stream of literature and the adaptive markets hypothesis. The adaptive markets hypothesis (AMH) stream of literature is an upcoming stream of literature in behavioral finance, which tries to bridge theory concerning the limits to arbitrage and theory about the efficient markets hypothesis. The AMH uses the evolutionary framework to explain limitations to arbitrage. We believe that using the context of automobile markets, we have successfully been able to test predictions relating to the existence of subgroups based on specific patterns of behavior. We are also able to show that behaviors change in response to both environmental factors and biases exhibited by arbitrageurs.

We study how arbitrageur specialization on two dimensions namely asset type and sourcing locations affect arbitrage profits. We categorize arbitrageurs into different groups based on various types of specialization behaviors. We model the arbitrage outcomes as a system of equations and try to understand how arbitrageur specialization affects the outcomes of arbitrage profits and arbitrage intensity, and vice-versa. These findings are important in light of the behavioral theory of finance, since they confirm certain long-held beliefs about arbitrageurs.

We believe that this dissertation based on spatial arbitrage will help researchers in information systems, economics and finance by providing a better understanding of the mechanisms behind spatial arbitrage and arbitrageur behavior. We believe that our contributions which include using spatial arbitrage as a measure of market efficiency and our understanding of arbitrageur evolution in the context of the adaptive markets hypothesis will be used by future generations of researchers.

Appendix A: For Chapter 2 - Procedures for estimating the transport cost between facilities

The intermediary that provided the data offers a transportation service by which a buyer can have vehicles transported from the intermediary's facilities to his location. Buyers used this service for approximately 35% of the purchased vehicles in our data. For those vehicles, our data contain the transport fee (*TransportFee*) between the facility and the buyer's location (which we know at the zip code level). We used these facility-to-buyer transport costs to estimate facility-to-facility transport costs. As our estimate of the transport cost between facilities k and l on day t, we used the mean of *TransportFee* for vehicles transported between facility k and buyer locations within x miles (based on zip code) of facility l – and vice versa – in month t. We set x=10 miles, and we aggregated to the month level to increase the stability of our estimates. In some cases, this did not yield an estimated transport cost between facilities k and l in month t due to lack of observations. We filled in these gaps by first increasing x to 30 and 50 miles and then extending the time period from month to quarter. We filled in remaining gaps for each facility-pair-month via linear interpolation. For example, if our transport cost estimates between Denver and Dallas for March, April, May, and June 2006 were null, $151, $152, and $153, we interpolated the missing value for March to be $150. This produced transport cost estimates for 99.85% of the possible arbitrage opportunities identified by the matching procedure described in §2.5.1.

Appendix B: For Chapter 2 - Handling Potential buyer heterogeneity

We considered whether buyer heterogeneity might confound our conclusions about the effects of the webcast channel and the standalone electronic market on spatial arbitrage. We discuss each channel in turn.

Potential buyer heterogeneity in the webcast channel analysis: We considered whether buyers who purchased in webcast enabled lanes might be different in unobservable ways from buyers who purchased in non-webcast enabled lanes. If they were, then this might be an alternative explanation for the negative effect of the webcast channel on spatial arbitrage. To explore this, we examined whether the buyers in webcast enabled lanes were a different set of buyers from those in non-webcast enabled lanes; this would be necessary for buyer unobservables to have a confounding effect. 95% of vehicles purchased from the physical market (by buyers participating either physically or via the webcast channel) between 2003 and 2007 (which is the time span for this part of our analysis) were purchased by 69,172 buyers. Of these buyers, 95% purchased in both webcast and non-webcast enabled lanes. In other words, there is little evidence of a distinct set of "webcast lane buyers" whose unobserved characteristics might differ from those of "non-webcast lane buyers"; they are (overwhelmingly) the same buyers.

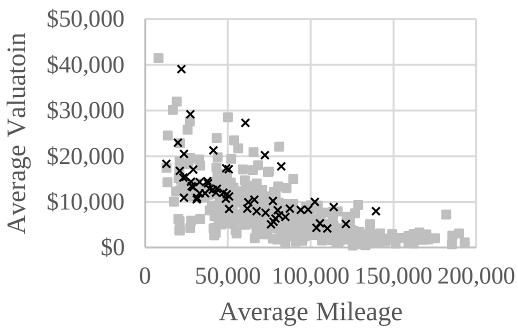
Of course, there is one major distinction between the buyers in webcast enabled lanes and those in non-webcast enabled lanes: the former are more likely to be from remote locations (as per the *RemoteBuyer* analysis reported in §2.5.2.3) and are likely more numerous (as per the *Price* analysis reported in §2.5.2.3). These differences in buyers explain *why* the webcast channel reduces spatial arbitrage; i.e., they are the mechanism behind the effect, as opposed to an alternative explanation for the effect. As we discuss in the main text, the webcast channel extends buyers' purchasing reach (i.e., the intermediate outcome), and this affects the probability that vehicles are arbitraged (i.e., the final outcome). If we attempted to control for this aspect of buyer heterogeneity, then we would be conditioning on an intermediate outcome. This is not appropriate in our setting and would bias the treatment effect (see Rosenbaum 1984).

To elaborate on this point, we compare our study to a hypothetical study designed to test the treatment effect of a drug for patients with dementia. In the dementia study, some patients receive the drug treatment, and some do not. In our study, some vehicles receive the webcast treatment, and some do not. In the dementia study, assume that the treated patients are significantly more likely to be able to follow a recipe compared to the non-treated patients. I.e., the drug has a treatment effect, and the likely mechanism for the effect is that the drug changes the composition of brain cells in the treated patients. In our study, the treated vehicles are significantly less likely to be arbitraged compared to the non-treated vehicles. I.e., the webcast channel has a treatment effect, and the likely mechanism for the effect is that the webcast channel changes the composition of buyers for the treated vehicles. There is no need to control for this type of post-treatment buyer heterogeneity in our study, just as there is no need to control for post-treatment "brain cell heterogeneity" in the dementia study.

Potential buyer heterogeneity in the standalone electronic market analysis: Similar to above, we examined whether the buyers who purchased in the standalone electronic market were a different set of buyers from those who purchased in the physical market (either physically or via the webcast channel). Although almost all of the buyers in the standalone electronic market also purchased via the physical market,[29] there is a large set of buyers who purchased in only the physical market. The "physical only" buyers differ from the "physical + electronic" buyers. On average, the latter purchase lower mileage, higher value vehicles, as illustrated in Figure A1. The two sets of buyers may be different on other dimensions as well. However, it is not clear why this heterogeneity would explain our result re: the

29 97% of vehicles purchased in the standalone electronic market were purchased by 21,206 buyers. Of these buyers, 98% purchased in both the standalone electronic and physical markets.

standalone electronic market's positive effect on arbitrage. Also, despite this heterogeneity, our results provide compelling evidence that the standalone electronic market fosters spatial arbitrage. Specifically, we show in §2.5.2.3 that the transaction immediacy of the standalone electronic market – which is not available in the physical market, either physically or via the webcast channel – enables arbitrageurs to identify and quickly purchase undervalued vehicles for later arbitrage. This improves arbitrageurs' ability to exploit arbitrage opportunities. We believe this provides a more compelling and plausible explanation for the positive effect of this market on spatial arbitrage than does the possibility of unobserved buyer heterogeneity.



Notes: The gray squares each represent a buyer who only purchased vehicles in the physical market. The black x's each represent a buyer who purchased vehicles in the physical market and in the standalone electronic market. The placement of each square reflects the average *Mileage* and *Valuation* of the vehicles purchased by a buyer. The chart is based on a random sample of buyers; using the full sample makes the graph too dense to interpret.

Figure B1: Scatter plot depicting the average Mileage *and* Valuation *of vehicles purchased by "physical only" buyers (gray squares) and "physical + electronic" buyers (black x's).*

We also implemented a robustness check to reduce the heterogeneity between the "physical only" and "physical + electronic" buyers in the standalone electronic market analysis. We reran the §2.5.2.2 analysis using only transactions for which *Mileage* was less than 54,411 and *Valuation* was greater than $13,420.30 (Note that Figure A1 shows that many "physical only" buyers purchased low mileage, high value vehicles similar to those purchased by "physical + electronic" buyers.) This eliminated "physical only" buyers who purchased very dissimilar vehicles from the analysis. Results are shown in Table B1

30 These are the mean values of *Mileage* and *Valuation* for the "physical + electronic" buyers. We used other thresholds and achieved similar results.

and are consistent with those in the main text. A related point is that by controlling for vehicle heterogeneity between the two markets in the matching analysis (by matching on *VehicleYear*, *Make*, *Model*, *Mileage*, and *Valuation*), we also control for unobserved buyer heterogeneity to some degree. This is because a used car dealer's characteristics are often defined by the vehicles he sells (e.g., a dealer might specialize in selling low-mileage German vehicles, in high-mileage U.S. vehicles, etc.) By comparing very similar vehicles across channels, we are likely comparing similar buyer populations. As further evidence that buyer heterogeneity is unlikely to confound our conclusion, the Rosenbaum sensitivity test reported in the main text shows that even if our result is partially attributable to unobservables such as buyer heterogeneity, then these unobservables would have to have a large effect (Γ=1.49) to overturn our conclusion. Last (but not least), many studies that compare markets find valid effects even though buyers are heterogeneous across markets (e.g., Dewan and Hsu 2004; Koppius and Van Heck 2004, Banker et al. 2011). For example, Dewan and Hsu (2004) compared two markets that differed in how much information they provided. They (validly) attributed the outcomes they observed to this information difference, even though buyers were heterogeneous across the two markets.

Table B1: Treatment effect of the vehicle being purchased in the standalone electronic channel on whether the vehicle is later arbitraged, using the sub-sample.

StandaloneElectronicMarket$_j$ (β_1)	0.541 (0.159) ***
Intercept (β_0)	-5.543 (0.123) ***
N	41,073
The dependent variable is the probability that the vehicle is later arbitraged. Model estimated via logistic regression. Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$). Results shown with α=7, where α denotes the number of days between flips used to delineate spatial arbitrage.	

References

Banker R, Mitra S, and Sambamurthy V (2011) The effects of digital trading platforms on commodity prices in agricultural supply chains. *MIS Quarterly* (35:3): 599-611.

Dewan S, and Hsu V (2004) Adverse selection in electronic markets: Evidence from online stamp auctions. *Journal of Industrial Economics* (52:4): 497-516.

Koppius O, van Heck E, and Wolters M (2004) The importance of product representation online: Empirical results and implications for electronic markets. *Decision Support Systems* (38:2): 161-169.

Rosenbaum P (1984) The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society: Series A*. 147(5): 656-666.

Appendix C: For Chapter 3 - Robustness tests

Table C1: Random effects Regression results with MeanProfits, NumArbitraged

	meanArbitrageProfit
VehicleSpecialization	0.0183***(0.0042)
LocationSpecialization	0.0078**(0.0033)
CapitalWk	-0.0003(0.0002)
NumCarsPurchased	0.0000**(0.0000)
proportionArbitrage	-0.0071(0.0069)
pctArbitragedeChannel	0.0190***(0.0041)
meanVehicleMileage	-0.0041***(0.0004)
totalOpportunities500	-0.0000(0.0000)
pcteAccess500	-0.0031(0.0206)
Year specific results	Included
_cons	0.0870***(0.0072)
N	4000
Degree Free. min	16.0000
Chi-squared	248.7320

Standard errors in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table C2: Regression results with MeanProfits, NumArbitraged Adjusted with Vehicle Make Specialization

	meanArbitrageProfit
VehicleSpecialization	0.0195***(0.0046)
LocationSpecialization	0.0075**(0.0035)
CapitalWk	-0.0000(0.0002)
NumCarsPurchased	0.0000(0.0000)
proportionArbitrage	-0.0064(0.0088)
pctArbitragedeChannel	0.0164***(0.0045)
meanVehicleMileage	-0.0039***(0.0006)
totalOpportunities500	0.0001*(0.0000)
pcteAccess500	0.0129(0.0229)
_cons	0.0740***(0.0092)
Year specific constants	Included
N	4000
R^2	0.601
adj. R^2	0.511
F-stat	8.4815
Degree Free. Min	16.0000
Degree Free. R	3264.0000
log-likelihood.	8787.0293

Standard errors in parentheses (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$)

Table C3: Regression results with Theil Index

	meanArbitrageProfit
VehicleSpecialization	0.0109***(0.0035)
LocationSpecialization	0.0073***(0.0028)
CapitalWk	-0.0001(0.0002)
NumCarsPurchased	0.0000(0.0000)
proportionArbitrage	0.0020(0.0085)
pctArbitragedeChannel	0.0172***(0.0045)
meanVehicleMileage	-0.0039***(0.0006)
totalOpportunities500	0.0001*(0.0000)
pcteAccess500	0.0132(0.0230)
Year	Included
_cons	0.0764***(0.0091)
N	4000
R^2	0.600
adj. R^2	0.509
F-stat	7.6721
Degree Free. min	16.0000
Degree Free. r	3264.0000
log-likelihood.	8779.3970
Chi-squared	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix D: For Chapter 3 - Clustering Algorithms for creating groups of arbitrageurs based on vehicle specialization and location specialization

We use the regression $VehicleSpecialization_{it} = \alpha_{0i} + \beta_{0i}.(t - 2000) + \beta_0^{'}(t\text{-}2000)^2$ and $LocationSpecialization_{it} = \alpha_{1i} + \beta_{1i}.(t - 2000) + \beta_1^{'}(t\text{-}2000)^2$ to obtain the coefficients (i.e $\alpha_{0i}, \beta_{0i}, \beta_0^{'}$ *and* $\alpha_{1i}, \beta_1, \beta_1^{'}$) for each arbitrageur *i*. We use the kmeans clustering algorithm to cluster groups of arbitrageurs based on their i.e $\alpha_{0i}, \beta_{0i}, \beta_0^{'}$ As a result, we obtain 3 groups of arbitrageurs G1,G2,G3. Similarly, we create 3 groups L1,L2,L3 of arbitrageurs based on $\alpha_{1i}, \beta_1, \beta_1^{'}$. We construct 9 groups of arbitrageurs such that each arbitrageur belongs to a either G1,G2,G3 or L1,L2,L3. The 9 groups and the mean trends of their specialization patterns are plotted in the below figure.

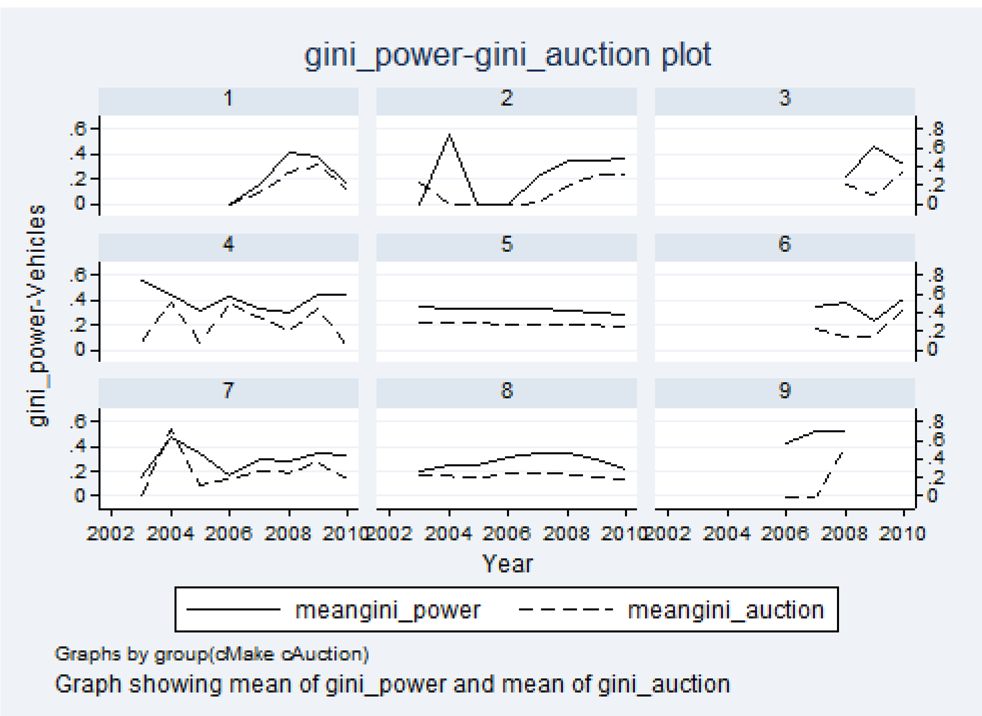


Figure D1: Trends showing mean Vehicle Specialization (Gini_power) and Location Specialization (Gini_auction)

Table D1: Arbitrageur count by group

group	1	2	3	4	5	6	7	8	9
count	11	8	2	5	547	4	16	109	1

I have also clustered on 6 variables namely (i.e α_{0i} ,β_{0i}, $\beta_0{}'$ and α_{1i} ,β_1, $\beta_1{}'$) to obtain 9 groups directly.

The main trends of VehicleSpecialization and LocationSpecialization are given in the diagram below.

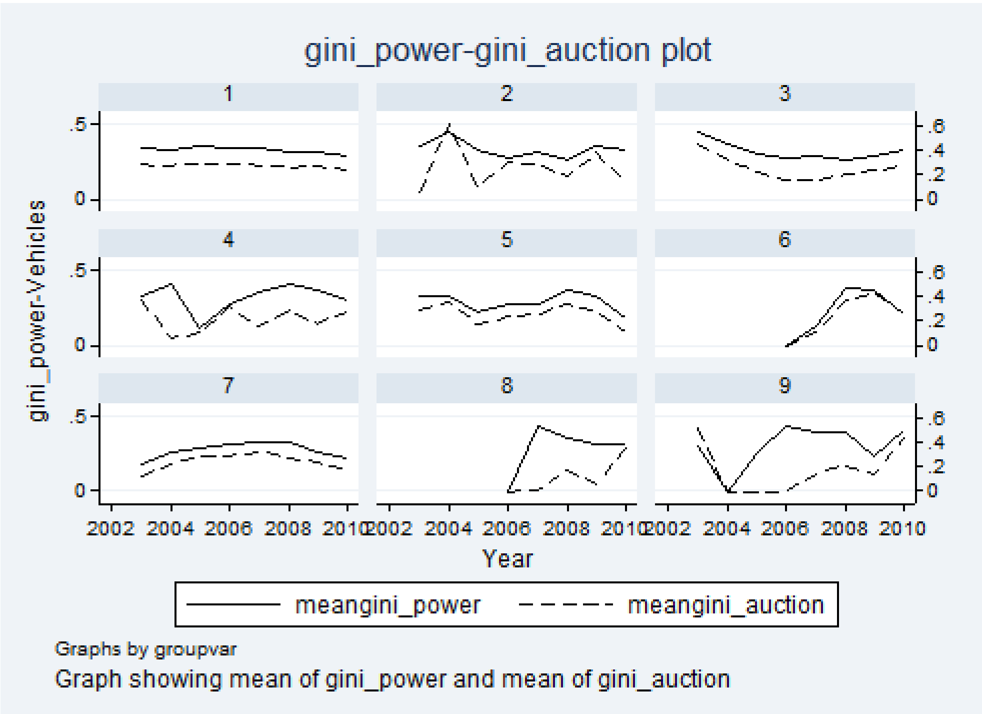


Figure D2: Trends showing mean Vehicle Specialization (Gini_power) and Location Specialization (Gini_auction) using coefficients of second order fits

Table D2: Arbitrageur count by group.

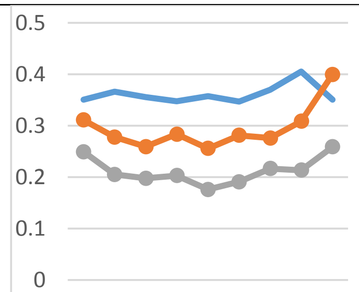
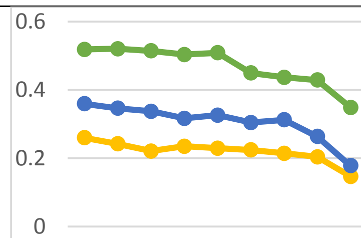
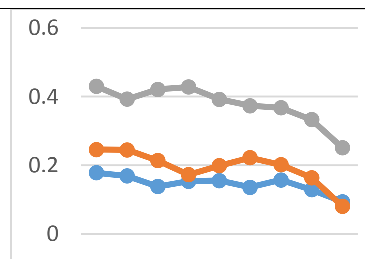
group	1	2	3	4	5	6	7	8	9
count	346	18	131	19	36	9	134	4	6

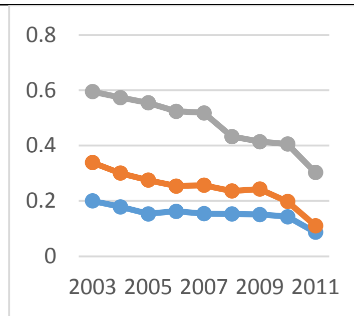
From the above figure D1 we see that the groups are not distributed evenly and the patterns cannot be uniquely discerned using the k-means clustering algorithm.

Appendix E: For Chapter 3 - Comparing different measures of specialization

The following table lists the important concentration measures we tested, and compared

Table E1: Different specialization measures and their mathematical formulations

Name of Measure	*Mathematical Formulation*	*Trends for the 3 groups of arbitrageurs for Vehicle makes.*	*Comments.*
HHI	$$HHI = \sum_{i=1}^{n} s_i^2$$		Extremely sensitive to the count. Does not exhibit uniform trends of specialization for the high – low and medium intensity arbitrageurs. Hence it is not a good measure and cannot accurately indicate vehicle specialization. The ranges are always 0 – 1. The use of HHI is mostly for industry concentration.
Gini	$G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert x_i - x_j\rvert}{2n^2\mu}$		Shows consistent trends for high – medium-low intensity arbitrageurs, that can be used for determining arbitrageur strategy. The trends are uniform for both vehicle specialization and auction specialization
Generalized Entropy	$GE(-1) = -\sum_{i=1}^{n} s_i \log s_i$		Extremely sensitive to count. The trends are similar to the Gini or Theil. However the range is much larger. Calculation is not straightforward
Mean Logarithmic standard deviation	$GE(0) = (\frac{1}{N})\sum_{i=1}^{N} \ln(\frac{\mu}{s_i})$	Similar to the above	Similar to Entropy above, except that the sensitivity is low to the bottom of the distribution.

Theil index	$GE(1) = (1/N)\sum_{i=1}^{N}(\frac{x_i}{\mu})\ln(\frac{x_i}{\mu})$		This measure follows uniform patterns for all 3 groups. It is robust and, though sensitive to Count can be used as an alternative measure to test our specialization results for robustness.
Half square of coefficient of variation	$GE(2) = \left(\frac{1}{2}\right)\left(\frac{\sigma}{\mu}\right)^2$	Similar to the above	Similar to the above, except the range is larger.
CR3	$CR(3) = \sum_{i=1}^{n} s_i$	N.A.	Is always 100 if there are only 3 types. Will not accurately indicate variations with respect to auction centers, where many arbitrageurs visit very few auction centers. Will have a skewed distributions if only a few different types exist.

Appendix F: For Chapter 3 – Results of mixed logit simulations

An alternative analysis for understanding factors affecting change in arbitrageur strategies

In order to understand factors affecting change in arbitrageur strategies, we construct a strategy profile for each arbitrageur, during a time period *t* between 2003 and 2010. The strategy profile consists of a series of states, with each state consisting of two variables, one for each dimension of specialization during a particular year. These variables indicate whether the arbitrageur was a high specialist type– denoted by alphabet *H* or low specialist type denoted by the alphabet *L*. If the arbitrageur didn't engage in arbitrage during a particular year the state would be denoted by *NN* for that particular year. We consider all values above the sample mean value of *H* and all values below the mean as *L* along both dimensions i.e. *VehicleSpecialization and LocationSpecialization*. An example strategy profile is given below in table 4 below:

Table F1: Sample strategy profile of an arbitrageur

Year	2003	2004	2005	2006	2007	2008	2009	2010
Profile	*LL*	*LH*	*HL*	*HH*	*HH*	*HH*	*HH*	*HH*

Each arbitrageur can be in one of the 5 possible states *i.e. HH, HL, LH, LL or NN* during a particular year. In order to understand factors that affect arbitrageur's likelihood of choosing a particular state s_t in a particular year *t* given that his previous state was s_{t-1} in year *t-1*, we use a mixed logit choice model (Train, 2009), which is a modification of the standard multinomial logit choice model. Each arbitrageur would choose to be in a state which maximizes his utility. The utility of each state *s* at time *t* to arbitrageur *i* is defined as follows:

$$U_{ist} = \alpha_{is} + \beta_s' x_{it} + \Sigma_{t=2}^{10} \gamma_{st} time_t + \Sigma_{m=1}^{5} \sigma_{im} d_{ms} + \epsilon_{ist}$$

Each term is described below in detail. See (Train, 2009) for more details about this. The probability that an arbitrageur *i* chooses state *s* is given by the following model.

$$p_{ist} = \frac{\exp\left(\alpha_{\text{is}}\ \beta_s' x_{it} + \Sigma_{\text{t=2004}}^{2010} \gamma_{st}\ time_t + \Sigma_{m=1}^{5}\ \sigma_{im} d_{ms} + \epsilon_{ist}\right)}{\Sigma_{s=1}^{S}\ \exp\left(\alpha_{\text{is}} +\ \beta_s' x_{it} + \Sigma_{\text{t=2004}}^{2010} \gamma_{st}\ time_t + \Sigma_{m=1}^{5}\ \sigma_{im} d_{ms} + \epsilon_{ist}\right)}$$

Here α_{is} is a normally distributed random intercept which captures the utility of each state s for each dealer *i*. α_{is} captures the unobserved preferences of each arbitrageur *i* across time periods. x_{it} is a vector which has variables that affect an arbitrageur's likelihood of choosing a new state. The variables constituting x_{it} are defined by: (i) $pctArbitraged_{it}$ which indicates the percentage of vehicles arbitraged by arbitrageur *i* in period *t*. Since most arbitrageurs are also traders, this variable i.e. $pctArbitraged_{it}$ indicates the importance accorded to arbitrage. (ii) $cntTotalVehicles_{it}$ indicates the total number of Vehicle purchased by the arbitrageur *i* from the market in a given year *t* irrespective of the vehicle being arbitraged or not. (iii) $pctElectronicVehicles_{it}$ denotes the total percentage of vehicles that were arbitraged and were sourced from electronic channels. (iv) $CapitalWk_{it}$ indicates the total amount of capital accessible per week by the arbitrageur *i* during the year *t.* (v) $totalAccessVehicles_{it}$ indicates the total vehicles accessible to the arbitrageur within 500 miles and (vi) $VehicleValuation_{it}$ which indicates the mean vehicle valuations of all vehicles purchased by arbitrageur *i*. The variable $time_t$ is included as a dummy to control for seasonality and unobserved time varying factors for each of the 7 periods (i.e. 2004-2010) of time in which we observe state transitions. The year 2003 is excluded because we do not observe the previous state. $\Sigma_{\text{t=2004}}^{2010} \gamma_{st}$ are values of associated coefficients. $\Sigma_{m=1}^{5}\ \sigma_{im} d_{ms}$ is included in the model, to account for the lack of independence amongst the error terms. The values of $d_{0s}..d_{4s}$ indicates the corresponding states (*HH, HL, LH, LL,NN*) and it is unlikely that error terms in our models are uncorrelated, on account of various factors such as a dealers preference for sourcing cars on electronic channels, or, a dealers preference for a particular location or vehicle type. This is referred to as a violation of the "IIA" assumption, as suggested in literature (Train 2009).We model this as a normally distributed random coefficient that varied across 5 corresponding states indicated by m = 1, 2, 3, 4 and 5. We also fitted a standard multinomial logit model and the alternative specifications constant logit model, which ignores dealer random effects and the violations of the IIA assumption. The results are similar and are displayed in Appendix F below.

The results of the mixed logit regressions are shown below in Table F4 and Table F5 respectively. Further discrete choice simulations are done by altering the 3 variables i.e *pctArbitraged*, *pctElectronicVehicles* and *totalCapitalWk.*

Table F2: Results of the mixed logit regression with 5 different models pertaining to originating states. These states are indicated by HH, HL,LL and NN. a1 is a dummy variable set to 1 if the state of arbitrageur i in period t is HH and set to 0 otherwise. Similarly a2 , a3 , a4 and a5 are dummy variables set to 1 for the corresponding states HL, LH, LL and NN respectively.

Previous state (t-1)	*HH*	*HL*	*LH*	*LL*	*NN*
	chosen	*chosen*	*chosen*	*Chosen*	*chosen*
Mean					
a1 x pctArbitraged		0.0205 (0.02)	13.09*** (3.63)	16.11*** (9.76)	17.41*** (7.60)
a2 x pctArbitraged	-0.926 (-1.18)		8.862*** (3.21)	15.81*** (9.57)	18.47*** (6.49)
a3 x pctArbitraged	-13.74*** (-9.37)	-9.329*** (-3.68)		11.28*** (7.83)	9.257*** (3.57)
a4 x pctArbitraged	-20.91*** (-10.96)	-12.40*** (-7.50)	-9.879*** (-2.97)		
a5 x pctArbitraged	-6.054 (-0.00)	-5.482 (-0.00)	-2.806 (-0.00)	9.907 (0.00)	-2.132 (-0.00)
a1 x cntTotalVehicles		0.00321** (2.64)	0.00481** (2.59)	0.00445*** (4.01)	0.00701*** (5.00)
a2 x cntTotalVehicles	-0.000774 (-1.17)		0.00154 (0.81)	0.00444*** (3.90)	0.00699*** (4.46)
a3 x cntTotalVehicles	-0.00591*** (-6.24)	-0.00445** (-2.20)		0.00309*** (2.90)	0.00581*** (2.83)
a4 x cntTotalVehicles	-0.00792*** (-7.78)	-0.00645*** (-4.67)	-0.00368** (-2.16)		
a5 x cntTotalVehicles	-0.00238 (-0.00)	-0.00243 (-0.00)	-0.00175 (-0.00)	0.00253 (0.00)	-0.00109 (-0.00)
a1 x pctElectronicVehicles		-10.37* (-1.91)	-2.375 (-0.34)	0.551 (0.15)	0.796 (0.13)
a2 x pctElectronicVehicles	2.701 (0.86)		9.915 (1.34)	-0.975 (-0.23)	-4.931 (-0.70)
a3 x pctElectronicVehicles	-1.692 (-0.50)	-3.336 (-0.41)		0.753 (0.23)	-12.22* (-1.75)
a4 x pctElectronicVehicles	7.725** (2.17)	-0.0807 (-0.02)	-4.598 (-0.99)		
a5 x pctElectronicVehicles	-0.862 (-0.00)	-2.486 (-0.00)	-3.869 (-0.00)	-0.215 (-0.00)	-3.637 (-0.00)
a1 x totalAccessVehicles		0.00383 (1.17)	0.00230 (0.46)	-0.000799 (-0.32)	-0.00423 (-1.04)
a2 x totalAccessVehicles	0.000385 (0.18)		0.00330 (0.65)	-0.0000822 (-0.03)	0.00413 (0.90)
a3 x totalAccessVehicles	-0.00126 (-0.55)	-0.00767 (-1.36)		-0.00307 (-1.38)	-0.00562 (-1.16)
a4 x totalAccessVehicles	0.000375 (0.16)	-0.00228 (-0.77)	-0.000280 (-0.09)		

Table F2 continued

a5 x totalAccessVehicles	-0.00852 (-0.00)	-0.00870 (-0.00)	-0.00504 (-0.00)	-0.00338 (-0.00)	-0.00844 (-0.00)
a1 x mnVehicleValuation		0.105** (2.06)	-0.0189 (-0.31)	0.0471 (1.28)	0.0132 (0.18)
a2 x mnVehicleValuation	-0.0538 (-1.59)		0.000212 (0.00)	0.0329 (0.83)	-0.0622 (-0.79)
a3 x mnVehicleValuation	-0.0867** (-2.53)	-0.0169 (-0.22)			
a4 x mnVehicleValuation	-0.0724** (-2.15)	-0.0466 (-1.00)	-0.0703 (-1.61)	0.0256 (0.75)	-0.0177 (-0.27)
a5 x mnVehicleValuation	-0.111 (-0.00)	-0.0928 (-0.00)	-0.0859 (-0.00)	0.0101 (0.00)	-0.137 (-0.00)
a1 x totalCapitalWk		-0.0983* (-1.71)	-0.0392 (-0.49)	0.00165 (0.03)	0.0168 (0.38)
a2 x totalCapitalWk	-0.0264 (-0.75)		-0.00420 (-0.04)	0.0160 (0.27)	0.00248 (0.05)
a3 x totalCapitalWk	0.0766* (1.82)	0.144 (1.52)		0.0176 (0.32)	-0.104 (-0.98)
a4 x totalCapitalWk	0.0873** (2.06)	0.128* (1.94)	0.112* (1.65)		
a5 x totalCapitalWk	0.00537 (0.00)	-0.0207 (-0.00)	0.0315 (0.00)	-0.00755 (-0.00)	-0.158 (-0.00)
alt== 2.0000	-2.692 (-0.83)		-12.33 (-1.59)	-3.519 (-0.82)	-0.961 (-0.00)
alt== 3.0000	4.555 (1.30)	3.516 (0.42)		-3.354 (-1.02)	0.275 (0.04)
alt== 4.0000	-3.244 (-0.89)	2.797 (0.58)	8.066 (1.63)		-4.433 (-0.01)
alt== 5.0000	-13.61 (-0.00)	-12.36 (-0.00)	-10.62 (-0.00)	-20.35 (-0.00)	
alt== 1.0000		7.841 (1.43)	0.0754 (0.01)	-4.851 (-1.30)	-14.22 (-0.00)
SD					
alt== 2.0000	-0.886*** (-4.51)		1.288** (2.13)	0.493 (0.87)	0.429 (0.33)
alt== 3.0000	-0.0987 (-0.22)	1.889** (2.47)		0.171 (0.21)	-0.907 (-0.81)
alt== 4.0000	0.533 (1.17)	0.0276 (0.05)	0.335 (0.29)		
alt== 5.0000	-0.165 (-0.00)	-0.0411 (-0.00)	-0.0450 (-0.00)	-0.116 (-0.00)	0.0335 (0.00)
alt== 1.0000		-1.338*** (-3.90)	-1.848*** (-3.00)	-0.634* (-1.83)	-0.0520 (-0.09)
time dummies[t2004-t2010] x state specific constants [a1..a5]	*Included*	*Included*	*Included*	*Included*	*Included*

Table F2 continued

Observations	6795	2645	2200	4190	1750
AIC	2477.5	1267.7	1074.7	1944.5	704.3
BIC	2859.7	1597.0	1393.7	2299.5	1005.0
log lik.	-1182.8	-577.8	-481.4	-916.2	-297.2
Chi-squared	484.8	112.1	59.30	237.3	126.3
t statistics in parentheses (* $p < 0.1$, ** $p < 0.05$, *** $p < 0.005$)					

Table F3: Results of the discrete choice simulation.

Previous State	*Current State*	*For each simulation, variables are increased by 1 standard deviation of their current values, while the rest of the variables are retained at their means. All values are in percentage(%).*					
@(t-1)	*@(t)*	*pctArbitraged*	*cntTotalVehicles*	*pctElectronicVehicles*	*totalAccessVehicles*	*mnVehicleValuation*	*totalCapitalWk*
HH	HH	20.92	21.90	-11.94	-0.18	4.40	-19.03
HH	HL	-	-	-	-	-	-
HH	LH	-100.00	-75.00	-100.00	-50.00	-50.00	1175.00
HH	LL	-95.02	-100.00	56.85	1.66	-19.50	68.46
HL	HH	21.20	161.96	-82.61	26.09	64.67	-78.26
HL	HL	49.21	-76.19	71.43	-13.76	-37.57	5.82
HL	LH	-	-	-	-	-	-
HL	LL	-84.62	-98.72	10.90	-14.10	-30.77	83.97
LH	HH	176.64	176.64	-17.76	4.67	9.35	-27.10
LH	HL	116.67	-83.33	916.67	16.67	33.33	-33.33
LH	LH	55.88	55.88	70.59	-2.94	94.12	-82.35
LH	LL	-73.38	-69.28	-20.48	-1.71	-15.02	20.14
LL	HH	180.38	113.29	-1.27	6.33	23.42	-9.49
LL	HL	0.00	0.00	0.00	200.00	0.00	200.00
LL	LH	97.73	81.82	20.45	-47.73	-75.00	56.82
LL	LL	-51.65	-33.86	-1.10	1.42	-0.63	-1.89
NN	HH	91.89	83.78	20.27	-22.97	12.16	9.45
NN	HL	-	-	-	-	-	-
NN	LH	25	425	-100.00	-75.00	0.00	-100.00
NN	LL	-41.26	-40.47	0.79	1.59	0.40	-0.79

Table F4. Results of Multinomial Logit

Previous state	1	2	3	4	5
1					
o.pctArbitraged		-0.840 (-1.03)	7.449*** (4.25)	15.32*** (10.94)	17.44*** (7.73)
o.cntTotalVehicles		0.00194** (2.32)	0.00276** (2.21)	0.00428*** (4.08)	0.00697*** (5.00)
o.pctElectronicVehicles		-7.637* (-1.87)	-0.908 (-0.19)	0.703 (0.21)	0.536 (0.09)
o.totalAccessVehicles		0.00296 (1.25)	0.00294 (0.84)	-0.000699 (-0.30)	-0.00428 (-1.05)
o.mnVehicleValuation		0.0794** (2.13)	-0.0244 (-0.56)	0.0212 (0.63)	0.0297 (0.67)
o.(mean) AggCapitalWk		-0.0639 (-1.51)	-0.0122 (-0.20)	0.00190 (0.04)	0.0180 (0.40)
o._cons		4.090 (1.35)	-1.176 (-0.33)	-3.918 (-1.51)	-4.814 (-1.01)
2					
pctArbitraged	-1.192* (-1.75)		6.851*** (3.43)	15.51*** (10.48)	18.25*** (7.70)
cntTotalVehicles	-0.000874 (-1.43)		0.000883 (0.55)	0.00443*** (4.05)	0.00691*** (4.58)
pctElectronicVehicles	2.330 (0.86)		6.413 (1.10)	-0.851 (-0.21)	-5.220 (-0.70)
totalAccessVehicles	0.000108 (0.06)		0.00366 (0.87)	-0.000118 (-0.04)	0.00390 (0.89)
mnVehicleValuation	-0.0451 (-1.52)		-0.0153 (-0.29)	0.00894 (0.24)	-0.0420 (-0.78)
(mean) AggCapitalWk	-0.0214 (-0.65)		0.0263 (0.34)	0.0118 (0.21)	0.00413 (0.08)
Constant	-1.804 (-0.89)			-3.453 (-1.14)	
o._cons		0 (.)	-6.758 (-1.54)		-1.831 (-0.33)
3					
pctArbitraged	-13.59*** (-9.41)	-7.858*** (-4.34)		11.30*** (8.05)	9.206*** (3.76)

Table F4 continued

cntTotalVehicles	-0.00591*** (-6.28)	-0.00364** (-2.45)		0.00316*** (2.97)	0.00551*** (3.00)
pctElectronicVehicles	-1.568 (-0.46)	-2.337 (-0.40)		0.644 (0.20)	-11.33* (-1.86)
totalAccessVehicles	-0.00122 (-0.54)	-0.00502 (-1.30)		-0.00308 (-1.40)	-0.00530 (-1.22)
mnVehicleValuation	-0.0877** (-2.58)	-0.0153 (-0.27)		-0.0251 (-0.74)	0.0174 (0.29)
(mean) AggCapitalWk	0.0787* (1.89)	0.119* (1.69)		0.0159 (0.29)	-0.0887 (-0.95)
o._cons			0 (.)	-2.666 (-1.10)	
Constant	3.721 (1.46)	2.423 (0.53)			5.526 (1.25)
4					
pctArbitraged	-20.35*** (-12.04)	-12.31*** (-7.67)	-10.10*** (-4.67)		
cntTotalVehicles	-0.00780*** (-8.04)	-0.00653*** (-4.88)	-0.00375** (-2.74)		
pctElectronicVehicles	7.346** (2.21)	0.419 (0.09)	-3.933 (-0.90)		
totalAccessVehicles	0.000466 (0.21)	-0.00184 (-0.64)	-0.000110 (-0.03)		
mnVehicleValuation	-0.0747** (-2.35)	-0.0493 (-1.09)	-0.0755* (-1.84)		
(mean) AggCapitalWk	0.0898** (2.19)	0.135** (2.13)	0.116* (1.87)		
o.pctArbitraged				0 (.)	0 (.)
o.cntTotalVehicles				0 (.)	0 (.)
o.pctElectronicVehicles				0 (.)	0 (.)
o.totalAccessVehicles				0 (.)	0 (.)

o.mnVehicleValuation				0 (.)	0 (.)
o.(mean) AggCapitalWk				0 (.)	0 (.)
Constant	-1.906 (-0.75)	2.274 (0.64)	5.166 (1.62)		
Observations	1359	529	440	838	350
AIC	2452.3	1249.2	1051.0	1911.5	672.7
BIC	2655.6	1415.8	1210.4	2096.0	823.1
log lik.	-1187.1	-585.6	-486.5	-916.7	-297.3
Chi-squared	630.2	191.3	162.0	254.3	214.1

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.005$

Table F5: Result of the ASCLogit regression

Previous state	(1)	(2)	(3)	(4)	(5)
alt					
a1pctArbitraged		-0.839 (-1.10)	7.449*** (4.25)	15.32*** (10.94)	17.45*** (7.73)
a2pctArbitraged	-1.192 (-1.64)		6.852*** (3.43)	15.51*** (10.48)	18.26*** (7.70)
a3pctArbitraged	-13.59*** (-7.83)	-7.857*** (-3.42)		11.30*** (8.05)	9.186*** (3.75)
a4pctArbitraged	-20.35*** (-5.96)	-12.31*** (-5.72)	-10.10*** (-4.67)		
a5pctArbitraged	-0.808*** (-3.17)	-2.655*** (-4.89)	1.823 (0.00)	5.659 (0.00)	6.117 (0.00)
a1cntTotalVehicles		0.00194** (2.59)	0.00276** (2.21)	0.00428*** (4.08)	0.00698*** (5.01)
a2cntTotalVehicles	-0.000874 (-1.03)		0.000884 (0.55)	0.00443*** (4.05)	0.00692*** (4.58)
a3cntTotalVehicles	-0.00591*** (-5.32)	-0.00364** (-2.22)		0.00316*** (2.97)	0.00550*** (3.00)
a4cntTotalVehicles	-0.00780*** (-5.55)	-0.00653*** (-2.82)	-0.00375** (-2.74)		
a5cntTotalVehicles	-0.000466** (-2.38)	-0.000973* (-1.72)	0.000179 (0.00)	0.00198 (0.00)	0.00317 (0.00)
a1pctElectronicVehicles		-7.636** (-2.02)	-0.907 (-0.19)	0.703 (0.21)	0.336 (0.05)
a2pctElectronicVehicles	2.330 (0.93)		6.412 (1.10)	-0.851 (-0.21)	-5.419 (-0.72)
a3pctElectronicVehicles	-1.568 (-0.48)	-2.337 (-0.37)		0.645 (0.20)	-11.99** (-1.99)

Table F5 continued

a4pctElectronicVehicles	7.346** (2.15)	0.419 (0.10)	-3.933 (-0.90)		
a5pctElectronicVehicles	0.332 (0.28)	-1.311 (-0.53)	0.604 (0.00)	-0.372 (-0.00)	-1.398 (-0.00)
a1totalAccessVehicles		0.00296 (1.25)	0.00294 (0.84)	-0.000698 (-0.30)	-0.00430 (-1.06)
a2totalAccessVehicles	0.000108 (0.06)		0.00366 (0.87)	-0.000118 (-0.04)	0.00388 (0.88)
a3totalAccessVehicles	-0.00122 (-0.57)	-0.00502 (-1.45)		-0.00308 (-1.40)	-0.00530 (-1.22)
a4totalAccessVehicles	0.000466 (0.22)	-0.00184 (-0.67)	-0.000110 (-0.03)		
a5totalAccessVehicles	0.0000112 (0.02)	-0.000208 (-0.12)	0.00148 (0.00)	-0.000683 (-0.00)	-0.00157 (-0.00)
a1mnVehicleValuation		0.0794** (2.41)	-0.0244 (-0.56)	0.0463 (1.35)	0.0136 (0.20)
a2mnVehicleValuation	-0.0451 (-1.32)		-0.0153 (-0.29)	0.0340 (0.90)	-0.0581 (-0.79)
a3mnVehicleValuation	-0.0877** (-2.29)	-0.0153 (-0.23)			
a4mnVehicleValuation	-0.0747** (-2.05)	-0.0493 (-0.90)	-0.0755* (-1.84)	0.0251 (0.74)	-0.0152 (-0.25)
a5mnVehicleValuation	-0.0176 (-1.60)	0.00620 (0.26)	-0.0215 (-0.00)	0.0405 (0.00)	0.0144 (0.00)
a1totalCapitalWk		-0.0639* (-1.69)	-0.0122 (-0.20)	0.00191 (0.04)	0.0183 (0.41)
a2totalCapitalWk	-0.0214 (-0.47)		0.0263 (0.34)	0.0118 (0.21)	0.00440 (0.08)
a3totalCapitalWk	0.0787* (1.66)	0.119 (1.51)		0.0159 (0.29)	-0.0867 (-0.94)
a4totalCapitalWk	0.0898* (1.78)	0.135 (1.45)	0.116* (1.87)		
a5totalCapitalWk	0.0157* (1.73)	0.0213 (0.73)	0.0232 (0.00)	-0.0191 (-0.00)	-0.150 (-0.00)
2					
Constant	-2.204 (-0.85)		-7.946 (-1.34)	-3.529 (-0.85)	0.954 (0.13)
3					
Constant	4.412 (1.29)	3.081 (0.47)		-3.244 (-0.99)	7.576 (1.30)
4					
Constant	-2.990 (-0.83)	2.303 (0.54)	7.463* (1.68)		

Table F5 continued

5					
Constant	-18.27*** (-15.01)	-16.88*** (-6.73)	-17.83 (-0.00)	-18.88 (-0.00)	-17.70 (-0.00)
1					
Constant		5.963 (1.55)	0.158 (0.03)	-4.821 (-1.37)	-4.112 (-0.63)
Year specific constants	Included	Included	Included	Included	Included
Observations	6795	2645	2200	4190	1750
AIC	2478.3	1275.2	1077.0	1937.5	697.4
BIC	2833.1	1581.0	1373.2	2267.2	976.3
log lik.	-1187.1	-585.6	-486.5	-916.7	-297.7
Chi-squared	215.8	120.4	106.6	164.2	110.5

t statistics in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.005$

REFERENCES

Aker JC (2010) Information from markets near and far: Mobile phones and agricultural markets in Niger. *American Economic Journal: Applied Economics*. 2(3): 46-59.

Alexander C, Wyeth J (1994) Cointegration and market integration: An application to the Indonesian rice market. *The Journal of Development Studies*. 30(2): 303-334.

Badiane O, Shively GE (1998) Spatial integration, transport costs, and the response of local prices to policy changes in Ghana. *Journal of Development Economics*. 56(2): 411-431.

Baffes J (1991) Some Further Evidence on the Law of One Price: The Law of One Price Still Holds *American Journal of Agricultural Economics* (73:4):1264-1273.

Bakos JY (1991) A strategic analysis of electronic marketplaces. *MIS Quarterly*. 15(3): 295-310.

Bakos Y (1998) The emerging role of electronic marketplaces on the Internet. *Communications of the ACM.* 41(8): 35-42.

Barberis N, Thaler R (2003) A survey of behavioral finance. *Handbook of the Economics of Finance* 1: 1053-1128.

Barrett C (2008) Spatial Market Integration. Durlauf S, Blume L, eds. *The New Palgrave Dictionary of Economics* (Palgrave Macmillan, London).

Baye MR, Morgan J, Scholten P (2006) Information, search, and price dispersion. Hendershott T, eds. *Handbook on Economics and Information Systems* (Elsevier, Amsterdam).

Blundell R, Bond S (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics.* 87(1): 115-143.

Brown JR, Goolsbee A (2002) Does the Internet make markets more competitive? Evidence from the life insurance industry. *Journal of Political Economy*. 110(3): 481-507.

Brynjolfsson E, Smith MD (2000) Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science*. 46(4): 563-585.

Case KE, Shiller RJ (1988) The Efficiency of the Market for Single-Family Homes. National Bureau of Economic Research. No. w2506.

Chellappa RK, Sin RG, Siddarth S (2011) Price formats as a source of price dispersion: A study of online and offline prices in the domestic U.S. airline markets. *Information Systems Research*. 22(1): 83-98.

Coleman A (2009) Storage, slow transport, and the law of one price: Theory with evidence from nineteenth-century U.S. corn markets. *The Review of Economics and Statistics*. 91(2): 332-350.

Curry B, George KD (1983) Industrial concentration: a survey. *The Journal of Industrial Economics*. 203-255.

De Jong A, Rosenthal L, Van Dijk MA (2009) The Risk and Return of Arbitrage in Dual-Listed Companies. *Review of Finance.* 13(3): 495-520.

Dimoka A, Hong Y, Pavlou PA (2012) On Product Uncertainty in Online Markets: Theory and Evidence. *MIS Quarterly*. 36(2): 395-426.

Fama EF (1970) Efficient capital markets: A review of theory and empirical work. *The Journal of Finance.* 25(2):383-417.

Fama EF (1998) Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics.* 49(3): 283-306.

Farmer JD (2002) Market force, ecology and evolution. *Industrial and Corporate Change.* 11(5): 895-953.

Froot KA, Dabora EM (1999) How are stock prices affected by the location of trade? *Journal of Financial Economics.* 53(2):189-216.

Gabaix X, Krishnamurthy A, Vigneron O (2007) Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market. *The Journal of Finance.* 62(2):557-595.

Gagnon L, Karolyi GA (2010) Multi-market trading and arbitrage. *Journal of Financial Economics.* 97(1):53-80.

Ghose A, Goldfarb A, Han SP (2013) How is the mobile Internet different? Search costs and local activities. *Information Systems Research*. 24(3): 613-631.

Ghose A, Yao Y (2011) Using transaction prices to re-examine price dispersion in electronic markets. *Information Systems Research*. 22(2): 269-288.

Grossman SJ, Stiglitz JE (1980) On the impossibility of informationally efficient markets. *The American Economic Review*, 393-408.

Hanson SG, Sunderam A (2014) The growth and limits of arbitrage: Evidence from short interest. *Review of Financial Studies.* 27(4): 1238-1286.

Hombert J, Thesmar D (2014) Overcoming limits of arbitrage: Theory and evidence. *Journal of Financial Economics.* 111(1): 26-44.

Iacus SM, King G, Porro G (2011) Causal inference without balance checking: Coarsened exact matching. *Political Analysis*. 20: 1-24.

Imbens G (2004) Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics*. 86(1): 4-29.

Jensen R (2007) The digital provide: Information (technology), market performance, and welfare in the south Indian fisheries sector. *Quarterly Journal of Economics*. 122(3): 897-924.

Kahneman D, Tversky A (1979) Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*. 263-291.

Kim JH, Shamsuddin A, Lim KP (2011) Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data. *Journal of Empirical Finance.* 18(5): 868-879.

Lintner J (1965) The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics.* 47(1):13-37.

Ljungqvist A, Qian W (2014) How constraining are limits to arbitrage? Evidence from a recent financial innovation. *National Bureau of Economic Research.*

Lo AW (2004) The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management.* 30(5):15-29.

Lo AW (2005) Reconciling efficient markets with behavioral finance: the adaptive markets hypothesis. *Journal of Investment Consulting. 7*(2): 21a-44.

Lung-Fei L (1992) Amemiya's generalized least squares and tests of overidentification in simultaneous equation models with qualitative or limited dependent variables. *Econometric Reviews* (11:3): 319-328.

McKenzie AM and Holt MT (2002) Market Efficiency in agricultural futures markets. *Applied Economics.* (34:12):1519-1532.

McMillan J (2002) *Reinventing the Bazaar: A Natural History of Markets*. (Norton & Co., New York).

McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*. 12: 153-157.

Mitchell M, Pulvino T (2001) Characteristics of risk and return in risk arbitrage. *The Journal of Finance.* 56(6): 2135-2175.

Mullainathan S, Thaler RH (2000) Behavioral economics. *National Bureau of Economic Research.*

Neely CJ, Weller PA, Ulrich JM (2009) The adaptive markets hypothesis: evidence from the foreign exchange market. *Journal of Financial and Quantitative Analysis.* 44(02): 467-488.

Norton E, Tenenbaum BH (1993) Specialization versus diversification as a venture capital investment strategy. *Journal of Business Venturing.* 8(5): 431-442.

Overby E, Clarke J (2012) A transaction-level analysis of spatial arbitrage: The role of habit, attention, and electronic trading. *Management Science*. 58(2): 394-412.

Overby E, Forman C (2015) The effect of electronic commerce on geographic purchasing patterns and price dispersion. *Management Science*. 61(2): 431-453.

Parker C, Ramdas K, Savva N (2013) Is IT enough? Evidence from a natural experiment in India's agriculture markets. Available at http://ssrn.com/abstract=2353771.

Peng L, Xiong W (2006) Investor attention, overconfidence and category learning. *Journal of Financial Economics*. 80(3): 563-602.

Persson KG (2008) The Law of One Price. Whaples R, eds. *EH.Net Encyclopedia of Economic and Business History* (Economic History Association, Tucson, AZ).

Pontiff J (1996) Costly arbitrage: Evidence from closed-end funds. *The Quarterly Journal of Economics.* 1135-1151.

Rosenbaum P (2002) *Observational Studies*. (Springer Science+Business Media, New York).

Sen M (2014) How judicial qualification ratings may disadvantage minority and female candidates. *Journal of Law and Courts*. 2(1): 33-65.

Sharpe W, Alexander G, Bailey J (1995) *Investments*. (Prentice Hall, Upper Saddle River, NJ).

Shleifer A (2000) *Inefficient markets: An introduction to behavioral finance.* (Oxford university press).

Shleifer A, Vishny R (1997) The Limits of Arbitrage. *Journal of Finance*. 52(1): 35-55.

Takayama T, Judge G (1971) *Spatial and Temporal Price Allocation Models*. (North-Holland, Amsterdam).

Takayama T, Judge G (1964) Equilibrium among spatially separated markets: A reformulation. *Econometrica: Journal of the Econometric Society*. 510-524.

Train KE (2009) *Discrete Choice Methods with Simulation*. (Cambridge University Press, Cambridge).

VITA

Hemang Subramanian enrolled in the IT Management Ph.D. program in 2010. Since early in the program, Hemang's research has dealt with key technology phenomenon affecting markets.
He employs econometric methods to analyze large volumes of data and to study how e-commerce influences spatial arbitrage in markets, and how arbitrageurs behave in markets. He also studies valuations of internet companies and how user generated content affects valuations.
He has presented his work at academic conferences including the CIST, INFORMS and at the Academy of Management doctoral consortium. One of his research articles is under second round revision at *Information Systems Research* and another of his research articles is about to be submitted for the third round revision at *Strategic Management Journal.* At Georgia Tech, he designed and successfully taught Business Programming course for 4 semesters, and the Continuing Education Project Management course. At the Scheller College of Business, he was nominated for the 2014 Ashford Watson Stalnaker Memorial Prize for Student Excellence in the Ph.D. program.

Before joining the Ph.D. program, Hemang worked at Yahoo! as an engineering manager and in several other engineering roles for 7 years. Prior to Yahoo!, Hemang worked at IBM Software labs as an engineer for about 4.5 years.

Hemang holds a B.E in Computer engineering from NITK, Surathkal, an M.S in Software Systems from BITS Pilani and the Executive Program in Business Management Certificate from the IIM Calcutta and is a Ph.D. Candidate in Information Technology Management at Georgia Institute of Technology.