Association for Information Systems
AIS Electronic Library (AISeL)

ICIS 2010 Proceedings

International Conference on Information Systems (ICIS)

2010

# DOES ELECTRONIC TRADING IMPROVE MARKET EFFICIENCY? EVIDENCE FROM SPATIAL ARBITRAGE IN THE AUTOMOTIVE MARKET

Eric Overby
*Georgia Institute of Technology - Main Campus*, eric.overby@mgt.gatech.edu

Jonathan Clarke
*Georgia Institute of Technology*, jonathan.clarke@mgt.gatech.edu

# DOES ELECTRONIC TRADING IMPROVE MARKET EFFICIENCY? EVIDENCE FROM SPATIAL ARBITRAGE IN THE AUTOMOTIVE MARKET

*Completed Research Paper*

**Eric Overby**
Georgia Institute of Technology
800 W. Peachtree St., Atlanta, GA 30308
eric.overby@mgt.gatech.edu

**Jonathan Clarke**
Georgia Institute of Technology
800 W. Peachtree St., Atlanta, GA 30308
jonathan.clarke@mgt.gatech.edu

## Abstract

*Price disparities across locations can occur when sellers in one location have difficulty matching with buyers in a different location due to the transaction costs of trading across distance. Spatial arbitrageurs exploit these discrepancies by buying goods from locations where prices are low and reselling them at locations where prices are high. Electronic channels should lower the transaction costs of trading across distance, thereby facilitating buyer/seller matching. It follows that electronic trading should reduce spatial arbitrage opportunities, thereby improving market efficiency. We test this hypothesis in the automotive market. The distinguishing feature of our data is that we can identify the distinct buyers, sellers, and vehicles involved in transactions, giving us a detailed look at transaction patterns likely motivated by spatial arbitrage. We conclude that traders are engaging in spatial arbitrage within the market but that spatial arbitrage has become less prevalent over time due to increased electronic trading.*

**Keywords:** Spatial arbitrage, electronic trading, market efficiency, automotive.

## Introduction

Supply and demand forces may cause essentially identical goods to trade at different prices in different geographic locations. This is the case in markets for goods such as automobiles, agricultural commodities, fuels, electricity, building materials, furniture, metals, industrial equipment and machinery, jewelry, and livestock (e.g., Coleman, 2009; Schuler, 2008). Different prices in different geographic locations create opportunities for traders to purchase a good at a location where prices are low, transport it to a location where prices are high, and resell it there at a profit. This behavior is referred to as *spatial arbitrage* (e.g., Baulch 1997; Fackler and Goodwin 2001.) Spatial arbitrage relates to the concept of *market integration*. An *integrated* market is one in which prices for the same goods at location $i$ equal prices at location $j$ plus the cost of transporting goods between locations $i$ and $j$. If a market is not integrated, then there are opportunities for spatial arbitrage.[1]

Spatial arbitrage arises due to transaction costs that prevent buyers and sellers in different locations from finding each other. For example, suppose seller $k$ at location $i$ has a product for which buyer $m_0$ at location $j$ is willing to pay $p_1$, but that seller $k$ and buyer $m_0$ are unable to match because they are in different locations. Suppose that this causes seller $k$ to settle for selling the product to buyer $m_a$ for price $p_0$, with $p_0 < p_1$. This creates a spatial arbitrage opportunity for buyer $m_a$, who can transport the product to location $j$ and sell it to buyer $m_0$ at a profit, assuming $p_1 - p_0$ exceeds the cost of transportation. Spatial arbitrage reveals market inefficiency, because two transactions are needed for buyer $m_0$ to purchase the product instead of one.

Electronic trading should reduce the transaction costs associated with trading across locations by making it easier for buyers and sellers to find and to conduct transactions with each other (Bakos 1998; Malone et al. 1987). This should make it easier for seller $k$ to transact with buyer $m_0$ directly, thereby eliminating buyer $m_a$'s spatial arbitrage opportunity and improving the efficiency with which buyer $m_0$ receives the product. The goal of the present study is to examine spatial arbitrage, market efficiency, and the effect of electronic trading in a specific setting – the wholesale automotive market.

We study this market for two reasons. First, we can observe spatial arbitrage with a high degree of precision. This is because each transaction contains buyer/seller ID's and the Vehicle Identification Number ("VIN"), which uniquely identify the buyer, seller, and vehicle. The granularity and detail of our data allow us to observe buyers who engage in spatial arbitrage by buying a vehicle at location $i$ for price $p_0$ and then reselling the same vehicle at location $j$ for price $p_1$. Second, the market has steadily been transitioning from physical trading to electronic trading. This allows us to examine the effect that electronic trading is having on spatial arbitrage.

The data consist of over 31 million transactions in the wholesale automotive market from January 1, 2003 to May 7, 2009. Transactions are conducted both physically and electronically, with the percentage of electronic transactions rising from approximately 1% in 2003 to approximately 18% in 2009. We identified 210,149 spatial arbitrage transactions, representing approximately 0.7% of the sample. We used logistic regression and discrete choice methods to examine the impact of electronic trading on spatial arbitrage. Results show that electronic trading activity is negatively associated with the likelihood of spatial arbitrage, thereby indicating that electronic trading increases market efficiency. There is also evidence that electronic trading forces arbitrageurs to move vehicles greater distances to maintain their profits. We also examined the reasons why some vehicles are arbitraged while others are not, concluding that the bounded rationality (due to limited attention) of the original sellers is a contributing factor.

The results contribute to two research streams: the stream on how electronic trading affects market efficiency and the stream on market integration and spatial arbitrage. First, the existing research on electronic trading and market efficiency has typically examined the phenomenon at a macro-level. For example, researchers have used price dispersion as a macro measure of market (in)efficiency and compared dispersion between online and offline markets to estimate the effect of electronic trading (e.g., Brynjolfsson and Smith 2000). By contrast, we examine the phenomena at a micro-level. We do this by using spatial arbitrage as the measure of market (in)efficiency and

[1] Spatial arbitrage should not be confused with arbitrage in a general sense, which typically refers to the simultaneous purchase and sale of the same asset for advantageously different prices (e.g., Sharpe and Alexander 1990.) Spatial arbitrage is a different, although related, concept. We will sometimes abbreviate "spatial arbitrage" as "arbitrage" for readability, but all references to "arbitrage" are specific to "spatial arbitrage."

examining how electronic trading activity affects the likelihood of spatial arbitrage on a transaction-by-transaction basis. The benefit of this approach is that it allows us to examine the specific mechanism through which electronic trading improves efficiency, which often is not possible at the macro-level. Second, the existing research on market integration and spatial arbitrage has been limited by the coarseness of available data. Barrett (2005) noted that studies in this stream often base conclusions solely on price data, despite the fact that the flow of goods between locations is critical to understanding the phenomenon. Data coarseness has also prevented researchers in this stream from investigating the behavioral factors that create arbitrage opportunities and the effect of electronic trading on arbitrage opportunities. The detail available in our data allows us to overcome this.

In the next section, we discuss the empirical setting for the study. We then develop hypotheses and describe the data used for hypothesis testing. We then present the analysis and results and conclude with a discussion of the contributions and limitations of the study.

## Empirical Context

The empirical context for the study is the wholesale automotive market, which is a business-to-business market for the exchange of used vehicles. Transactions in this market in the United States have traditionally occurred at physical market facilities where buyers, sellers, and vehicles are collocated.[2] Vehicles are traded at physical market facilities located throughout the United States. There are multiple intermediaries, referred to as automotive auction companies, that operate these facilities. Sellers transport vehicles to market facilities, where they are auctioned in a sequential format in which each vehicle is driven, one at a time, into the midst of a group of potential buyers. An auctioneer solicits bids for each vehicle, and the seller has the option to accept or reject the high bid. In the past 10-15 years, electronic trading channels have been added to the market, which we discuss in more detail below.

Buyers in the wholesale automotive market are used car dealers who use the market to procure approximately 35% of the vehicles they sell to retail consumers.[3] As this is a wholesale market, retail consumers are not allowed to purchase vehicles, with the exception of vehicles that have open access requirements (typically government vehicles.) Although most dealers purchase vehicles in the market for the purpose of reselling them to retail customers, some dealers purchase vehicles in order to engage in spatial arbitrage. For example, these dealers may purchase a vehicle for price $p$ at market facility $i$, transport it to market facility $j$, and resell it there for price $p^*$, with the expectation that $p^* > p$. We refer to these dealers as *spatial arbitrageurs*. We refer to the location at which a spatial arbitrageur buys a vehicle as the *source* location and the location at which the arbitrageur sells the vehicle as the *destination* location.

Sellers in the market are firms who own multiple vehicles, including rental car companies (e.g., Hertz), banks (e.g., Bank of America), automotive manufacturers (e.g., Toyota, Ford) and their finance arms (e.g., Toyota Financial Services, Ford Credit), and government agencies (e.g., police departments.) These firms choose to sell in the wholesale market for several reasons. First, many lack retail outlets and thus are not equipped to sell to the general public. Second, the wholesale market is generally more liquid and predictable than the retail market, which allows sellers to dispose of multiple vehicles quickly at a predictable price, although wholesale prices are typically lower than retail prices. Used car dealers may also be sellers in the market, in addition to their role as buyers. Perhaps the most common reason for a dealer to be a seller in the wholesale market is if s/he cannot or chooses not to sell a vehicle in the retail market. In this case, the dealer sells wholesale to another dealer who then attempts to retail the vehicle. Another reason for a dealer to be a seller is if s/he is engaged in spatial arbitrage.

---

[2] The market operates differently in other parts of the world. For example, see Lee and colleagues (1998; 1999) for analyses of the wholesale automotive market in Japan.

[3] Used car dealers obtain about 50% of the vehicles they sell as trade-ins and the other 15% in miscellaneous ways. Source: NADA Data 2009 (www.nada.org/nadadata), page 10. NADA stands for National Automobile Dealers Association.

## Hypotheses Development

### *Hypothesis #1: Existence of Spatial Arbitrage*

It is common for prices for the same product to vary across geography. This type of price dispersion has been documented repeatedly (see Baye et al. 2006 for a review). In another review, Fackler and Goodwin (2001) suggested that spatial arbitrage opportunities are rampant in markets for agricultural commodities but are often not exploited due to the cost of transporting products relative to their total value. In our case, we expect prices for the same or essentially similar vehicles to differ across geography, as they do in markets for other types of products. We also expect these differences to be larger than the costs of transportation in many cases. This is because the average price of vehicles in our sample is approximately $10,000, such that even a small percentage difference in price is a non-trivial raw amount. Thus, we expect to see some amount of spatial arbitrage in the market. This hypothesis is fundamental to the paper and is stated formally below.

*H1: Spatial arbitrage opportunities exist and are exploited in the wholesale automotive market.*

### *Hypothesis #2: Limited Attention and Spatial Arbitrage*

Conditional on support for H1, our second hypothesis relates to why spatial arbitrage opportunities occur. The existence of spatial arbitrage indicates that sellers are making suboptimal choices of where to sell some of their vehicles and that arbitrageurs are taking advantage of this by purchasing these vehicles and reselling them at higher-priced locations. A natural question is why sellers do not sell these vehicles initially at these higher-priced locations, thereby retaining the arbitrageurs' profits for themselves. We posit that one reason that sellers are leaving this money "on the table" is bounded rationality due to limited attention.

Building on the seminal work of Simon (1955), a large body of literature argues that decision makers are limited in their ability to process information and to perform multiple tasks simultaneously. Notably, Kahneman (1973) suggested that an individual's attention spent on one task must necessarily reduce the attention available for other tasks because of limited processing capabilities. Researchers have recently begun to assess how the limited attention of market participants impacts market efficiency. Corwin and Coughenour (2008) found evidence that specialists on the New York Stock Exchange allocate their limited attention to their most active stocks during periods of increased activity, leading to higher transaction costs for their remaining stocks. Peng and Xiong (2006) showed that investors resort to categorical learning behavior and process more market-wide information at the expense of firm-specific information because of limited attention.

In our context, sellers have multiple vehicles in their inventory and must attempt to choose the optimal selling location for each of them. We posit that sellers will pay significant attention to the selling location of vehicle models which comprise a large portion of their inventory but limited attention to vehicle models which do not. This parallels the limited attention arguments used by Corwin and Coughnour (2008) for NYSE specialists. This limited attention to vehicle models that make up a small portion of a seller's inventory will make sellers more likely to choose sub-optimal selling locations for vehicles of these models. This will increase the probability that these vehicles will be arbitraged. This leads to hypothesis 2.

*H2: Seller inattention to a given vehicle model is positively associated with spatial arbitrage activity on vehicles of that model.*

### *Hypothesis #3: Electronic Trading and Incidence of Spatial Arbitrage*

As discussed above, the U.S. wholesale automotive market has traditionally operated as a physical market. In the past 10-15 years, the automotive auction companies have introduced electronic channels into the market, although these channels have only accounted for a significant portion of transaction volume in the last few years. The most commonly used electronic channel is the webcast channel. This channel operates by simulcasting via the Internet the auctions as they are occurring at the physical facilities. Buyers can log into the webcast from an Internet browser and place bids in competition with buyers who are physically present at the facility.

We hypothesize that the webcast channel will reduce spatial arbitrage in the market. The intuition behind this hypothesis is that the webcast channel makes it easier for buyers to shift their demand to different locations. For example, buyers in locations where prices are high may shift their demand to locations where prices are low. Essentially, this means that any buyer can use the webcast channel to source a vehicle from a low-priced location and move it to a high-priced location, a role which was previously reserved for spatial arbitrageurs. Thus, we expect webcast use to be associated with reduced spatial arbitrage. We use the following model to present this argument more formally. We begin with a baseline scenario in which electronic trading is not available. We then examine the effect of introducing the webcast channel.

**Baseline Scenario**

First, note that the price discovery mechanism in our context is an auction. Assume that there are $n$ bidders competing for vehicle $g$ being auctioned by seller $k$ at location $i$ at time $t$. Each bidder $m$ has a valuation $v$ for vehicle $g$. Let $V_g$ be the array of bidder valuations for vehicle $g$, ordered from lowest to highest, i.e., $V_g \in \{v_{(1)}, v_{(2)}, \ldots, v_{(n-1)}, v_{(n)}\}$. Denote the bidder with the highest valuation (i.e., $v_{(n)}$) as bidder $m_a$. Based on standard auction theory pertaining to ascending auctions (Milgrom & Weber, 1982), the high bid is equal to the second highest valuation in $V_g$, i.e., $v_{(n-1)}$. Thus, the outcome of the auction is that bidder $m_a$ purchases vehicle $g$ at price $v_{(n-1)}$, assuming $v_{(n-1)}$ exceeds seller $k$'s reserve price.

Next, assume that there are bidders $m_1$ and $m_2$ at location $j$ who value vehicle $g$ at $v_{g,m1}$ and $v_{g,m2}$, respectively, where $v_{g,m1} > v_{(n-1)}$ and $v_{g,m2} > v_{(n-1)}$, but that these bidders do not bid for vehicle $g$ because they are in the wrong location. The existence of bidders $m_1$ and $m_2$ presents bidder $m_a$ with a spatial arbitrage opportunity. To wit, bidder $m_a$ can move the vehicle to location $j$ and resell the vehicle at time $t+1$ at price $v_{g,m1}$ or $v_{g,m2}$, whichever is smaller (because the winner of this auction need only pay the valuation of the runner-up.) If the difference between $v_{g,m1}$ or $v_{g,m2}$ and $v_{(n-1)}$ exceeds the cost of transporting the vehicle, then it is rational for bidder $m_a$ to behave this way.[4]

**Effect of Introducing the Webcast Channel**

The webcast channel allows buyers to participate in auctions at remote locations. Assume that bidder $m_1$ (or bidder $m_2$ or both bidders – the model holds regardless) uses the electronic channel to participate in the original auction at location $i$. Assume that s/he adjusts his/her valuation downward to account for the cost of transporting the vehicle from location $i$ to location $j$, denoted $c^T_{i,j}$. Thus, rewrite his/her valuation as $v_{g,m1}* = v_{g,m1} - c^T_{i,j}$. If $v_{g,m1}* \leq v_{(n-1)}$, then bidder $m_1$'s participation affects neither the outcome of the auction nor bidder $m_a$'s spatial arbitrage opportunity. However, if $v_{g,m1}* > v_{(n-1)}$, then bidder $m_1$'s participation shifts the array of bidder valuations from $V_g \in \{v_{(1)}, v_{(2)}, \ldots, v_{(n-1)}, v_{(n)}\}$ to either $V_g \in \{v_{(1)}, v_{(2)}, \ldots, v_{(n-1)}, v_{g,m1}*, v_{(n)}\}$ or $V_g \in \{v_{(1)}, v_{(2)}, \ldots, v_{(n-1)}, v_{(n)}, v_{g,m1}*\}$, depending on whether $v_{g,m1}* > v_{(n)}$. Thus, the high bid would increase from $v_{(n-1)}$ to either $v_{g,m1}*$ or $v_{(n)}$, whichever was lower. This means that bidder $m_1$ would either be: a) the second highest bidder and force bidder $m_a$ to pay $v_{g,m1}*$, or b) the high bidder and pay $v_{(n)}$. This limits the spatial arbitrage opportunity available to bidder $m_a$. In the first case, bidder $m_a$ must pay more for the vehicle. This reduces his/her potential spatial arbitrage profit by the difference between $v_{g,m1}*$ and $v_{(n-1)}$, thereby lowering the probability that s/he will engage in arbitrage. In the second case, bidder $m_a$ will not win the vehicle, precluding any arbitrage opportunity. This leads to hypothesis 3.

*H3: Webcast bidding activity that affects the price of a vehicle is negatively associated with spatial arbitrage activity on that vehicle.*[5]

### *Hypothesis #4: Electronic Trading and Conduct of Spatial Arbitrage*

We also posit that the webcast channel will influence the arbitrageur's choice of destination location. To motivate this hypothesis, we first assume that arbitrageurs: a) prefer to purchase vehicles in low-priced locations and resell

[4] There are additional transaction costs that bidder $m_a$ (and all arbitrageurs) must consider, including transaction fees, the opportunity cost of capital, etc. Discussion of these is withheld due to space limitations but is available from the authors.

[5] As assumption underlying H3 is that webcast bidders who affect a vehicle's price would not have otherwise been physical bidders if the webcast channel wasn't available. We provide evidence of the validity of this assumption below.

them in high-priced locations, and b) prefer to resell at locations close to the source location, all else equal. Similarly, we assume that buyers who use the webcast channel: a) are using the channel to source vehicles at prices lower than those at their location, and b) prefer to purchase at locations close to their location, all else equal. This means that the most attractive destination locations for the spatial arbitrageur are also likely to be the locations of the webcast bidders who place bids at the source location. This will make these locations less attractive destinations for spatial arbitrageurs. To illustrate, assume that pickup trucks trade at low prices in Dallas but at high prices in Houston. This will create an incentive for spatial arbitrageurs to purchase trucks in Dallas and to resell them in Houston, but it will also create an incentive for buyers located in Houston to use the webcast channel to purchase trucks directly from Dallas. This will cause Houston-based buyers to lower their valuations for trucks sold in Houston, thereby making Houston a less attractive destination location for the spatial arbitrageurs. Because webcast bids are likely to be from bidders in nearby locations, we posit that when webcast bidding affects the price of a vehicle, speculators will choose to transport vehicles to more distant destination locations to complete the spatial arbitrage. This leads to hypothesis 4.

*H4: Webcast bidding activity that affects the price of a vehicle is positively associated with the distance that spatial arbitrageurs move the vehicle to complete the spatial arbitrage.*

## Data

The data consist of all completed transactions (n = 31,805,961) facilitated by a major automotive auction company between January 1, 2003 and May 7, 2009. For each transaction, the data contain identification numbers for the buyer and seller (*BuyerID* and *SellerID*), the Vehicle Identification Number (*VIN*), the transaction date (*Date*), the transaction price (*Price*), the vehicle's make and model (*VehicleModel*), the vehicle's odometer reading (*Mileage*), the location at which the transaction occurred (*LocationID*), the vehicle's estimated wholesale valuation (*Valuation*), and an indication of the vehicle's history prior to being sold in the wholesale market (*SourceType*). There are 93 locations in the data, which are distributed throughout the United States. *Valuation* is calculated by the auction company based on transactions for similar vehicles over the prior 30 days. *SourceType* takes one of five values, depending on the type of seller and how the vehicle was used in the past. These values are: 1) factory, which refers to vehicles sold by automotive manufacturers after having been used as company cars, 2) lease, which refers to vehicles sold by leasing companies after the lease has expired, 3) rental, which refers to vehicles sold by rental car firms after the vehicle has been retired from rental service, 4) repossession, which refers to vehicles sold by financial institutions after the vehicle has been repossessed, and 5) dealer, which refers to vehicles sold by dealers.

The data also contain the channel through which the buyer purchased the vehicle. Buyers can purchase vehicles either in person via the physical channel or via the webcast channel (as discussed above). We use a dummy variable (*Webcast_Buyer*) to denote whether a vehicle was purchased by a buyer using the webcast channel. Note that the webcast channel does not affect the mechanism used to determine prices. Prices are determined via an ascending auction; a human auctioneer solicits bids until the highest bid is registered. The webcast channel simply gives bidders the option to place bids electronically if they cannot (or choose not to) travel to the market facility. Also, the webcast channel does not require a behavioral change by the seller, because the seller presents his/her vehicles the same way (having them driven into the physical facility) regardless of whether bidders are placing bids in person or via the webcast.[6]

For transactions in which the buyer purchased the vehicle using the physical channel, the auction company records whether the second-highest bid was placed by a bidder using the webcast channel. We coded this as a dummy variable (*SecondHighBid_Webcast*). The auction company records this data to help quantify the impact of the

[6] There is another electronic channel available, which is a traditional electronic market in which each vehicle is described on a web page where bidders can place bids or purchase the vehicle outright at a posted price. Transactions in this channel, which we refer to as the standalone electronic channel, represent 1.2% of the sample. The major distinction between this channel and the webcast channel is that the webcast channel augments the physical channel, while the standalone electronic channel is distinct from it. We dropped standalone electronic channel transactions from our analysis because the price discovery mechanism in this channel differs from that used in the physical and webcast channels. Dropping these transactions allows us to maintain correspondence between the theoretical model and the empirical context.

webcast technology even if a webcast bidder does not win the auction. If a webcast bidder wins the auction (i.e., *Webcast_Buyer* = 1), this data is not recorded. In that case, we set *SecondHighBid_Webcast* = 0.

We constructed variables from other variables in the data, including our measure of sellers' limited attention. We measured seller *in*attention to a given vehicle model *h* as an inverse function of the number of vehicles of other models in seller *k*'s inventory at time *t*. We based this measure on the inattention measures developed by Corwin and Coughenour (2008). To construct this measure, we first counted the total number of vehicles offered by seller *k* at time *t*, irrespective of location ($TotalVehicles\_Seller_{k,t}$). We then counted the total number of vehicles of a specific model *h* offered by seller *k* at time *t*, irrespective of location ($TotalVehicles\_SellerModel_{h,k,t}$). We defined $SellerInattentionProxy_{h,k,t} = (TotalVehicles\_Seller_{k,t} - TotalVehicles\_SellerModel_{h,k,t}) / TotalVehicles\_Seller_{k,t}$. For example, assume seller *k* offered 10 vehicles on day *t*, 9 of which were Nissan Maximas and 1 of which was a Nissan Pathfinder. For the Nissan Maximas, $SellerInattentionProxy_{h,k,t}$ would be (10 – 9 ) / 10 = 0.1, and for the Nissan Pathfinder, $SellerInattentionProxy_{h,k,t}$ would be (10 – 1 ) / 10 = 0.9. The intuition behind this measure is that seller *k* will allocate more attention to optimizing the selling location for Maximas than for Pathfinders, because Maximas make up a larger percentage of the seller's inventory. Thus, the level of our inattention measure should be higher for Pathfinders than for Maximas.[7] Our label for the measure indicates that it is a proxy for seller inattention, as it would be impossible to measure the actual level of inattention that each seller paid each vehicle over the 6+ years of the sample.

## Analysis and Results

### *Delineation of Speculative Activity*

Although spatial arbitrage has been studied in prior research (see the Introduction section), it has been difficult to observe spatial arbitrage behavior with precision because most data sets do not identify the individual buyers, sellers, and assets involved in transactions. In other words, most data sets do not permit the analyst to examine whether a specific trader purchased a specific item at location *i* at time *t* and then resold that same item at location *j* at time *t+1*. Because the *BuyerID*, *SellerID*, and *VIN* variables in our data uniquely identify buyers, sellers, and vehicles, we observe vehicles being flipped in this manner. However, we must infer whether these flips are due to spatial arbitrage or to other factors.

In our context, a *flip* is a pair of transactions for the same vehicle (identified by its VIN) in which the buyer in the first transaction is the seller in the second transaction. A dealer may flip a vehicle for several reasons. First, a dealer may flip a vehicle because s/he is engaging in spatial arbitrage. Second, a dealer may flip a vehicle after improving it (e.g., repairing dents, painting, etc.), with the expectation that the profits from the flip will exceed the cost of improvement. Third, a dealer may flip a vehicle if s/he is unable to sell it in the retail market and chooses to liquidate it via the wholesale market. Fourth, a dealer may flip a vehicle if s/he sells it to a retail customer and then regains possession at a later date, perhaps if the retail customer trades in the vehicle as part of a subsequent transaction or defaults on his/her loan. There are 2,123,718 flips in the data.

We focus on *cross-location flips*, or flips in which the vehicle is moved from the location at which it was purchased to a different location for resale. Not only is this appropriate based on the nature of our research questions and hypotheses, but it also helps us differentiate spatial arbitrage from other types of flips. For example, we cannot tell from our data if the buyer performing the flip improved the vehicle. However, we assume that if s/he did, s/he would resell the vehicle at the same location from which s/he purchased it in order to eliminate the cost of transporting the vehicle to another location. We make the same assumption for flips due to failure to retail the vehicle and vehicle repossession. Thus, by studying only cross-location flips, we reduce the possibility that the flips that we consider to be motivated by spatial arbitrage are better explained by other reasons. We also used *DaysToFlip*, which is the number of days between the two transactions comprising a flip, to delineate spatial arbitrage from other types of flips. We reasoned that spatial arbitrage flips would be completed the most quickly, i.e., they would have the lowest *DaysToFlip*. This is because the goal of the spatial arbitrageur is to maximize profits, and each day that s/he retains ownership of a vehicle reduces his/her profits due to his/her cost of capital and depreciation. Flips due to improving

[7] A vehicle's value might also affect how much attention the seller pays it, which we control for in our empirical analysis by including *Valuation* in the specification.

the vehicle would be completed the second-most quickly. This is because the buyer will still wish to minimize the number of days s/he retains the vehicle, but s/he must retain the vehicle for some period of time to complete the improvements. Flips due to failure to sell in the retail market are likely to be completed more slowly to account for the time that the buyer is attempting to retail the vehicle. Last, flips due to re-ownership of a previously retailed vehicle should take the longest time to complete, with potentially years passing between the original purchase and the subsequent resale in the wholesale market.

Figure 1 depicts the count and average gain (loss) of cross-location flips completed within 60 days. The y-axis represents *Flipped_PriceDifference*, which is the price in the first transaction subtracted from the price in the second transaction. The x-axis represents *DaysToFlip*, and the size of the bubbles represents the number of flips per *Flipped_PriceDifference* level. Two patterns are noteworthy. First, the number of flips declines with *DaysToFlip*, peaking at *DaysToFlip* = 6.[8] Second, the average gain declines with *DaysToFlip*. This can be seen visually by noticing that more of the mass of the bubbles is below 0 at higher *DaysToFlip*. Also, the correlation between average *Flipped_PriceDifference* and *DaysToFlip* shown in Figure 1 is −0.77. For comparative purposes, the means of *Flipped_PriceDifference* for *DaysToFlip* = 7, 21, 35, and 49 days are \$1425, \$1372, \$1216, and \$868 respectively. The average transaction price for the full sample is \$10,281, which helps place the values of *Flipped_PriceDifference* in context.

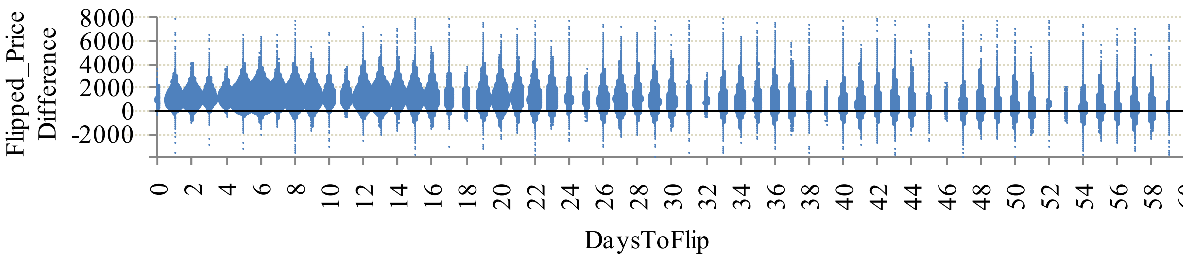


**Figure 1: *Flipped_PriceDifference* by *DaysToFlip*. Bubble size reflects the number of flips at that gain/loss for that day.**

Based on our reasoning and the descriptive data depicted in Figure 1, we classified a cross-location flip as spatial arbitrage if it was completed within $\alpha$ days of the original purchase, i.e., if *DaysToFlip* $\leq \alpha$. This assignment procedure is likely to yield false positives (e.g., categorizing a flip as arbitrage when it should not be) if $\alpha$ is set too high and false negatives (e.g., failing to categorize a flip as arbitrage when it should be) if $\alpha$ is set too low. In our hypothesis tests, we set $\alpha$ = 7, and we varied this threshold up and down to test for sensitivity. We used a conservative (i.e., low) $\alpha$ threshold for our main results to limit the possibility that cross-location flips were motivated by a factor other than arbitrage. There are 210,149 cross-location flips within the data for $\alpha$=7. Figure 2 shows the number of cross-location flips for different *DaysToFlip*. The cumulative line represents the total cross-location flips up to and including *DaysToFlip*; this represents the number of observations classified as spatial arbitrage flips at different levels of $\alpha$.

---

[8] The vast majority of transactions at each location are conducted on the same day each week (referred to as "sale day" within the industry), with the specific day depending on the location. For example, sale day for locations *i*, *j*, and *k* may be Tuesday, Thursday, and Monday, respectively. Thus, the most common values of *DaysToFlip* for flips between locations *i* and *j* will be 2 and 9, while the most common values of *DaysToFlip* for flips between locations *i* and *k* will be 6 and 13.

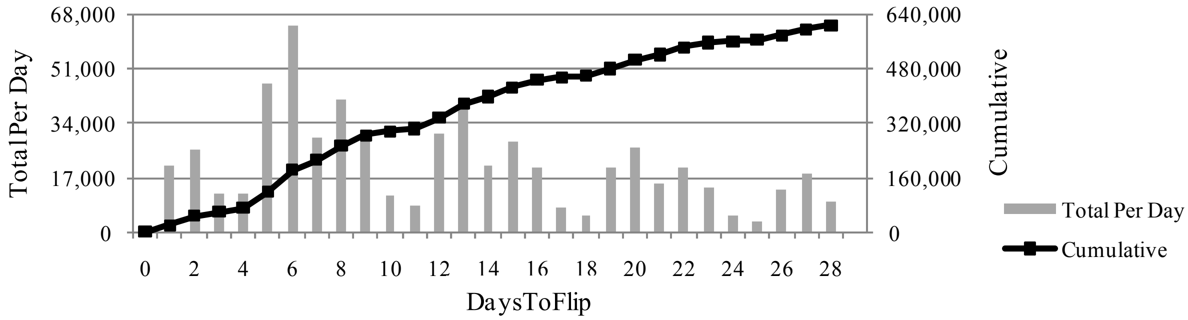


**Figure 2: Number of cross-location flips per *DaysToFlip*. Cumulative line represents the total cross-location flips up to and including *DaysToFlip*.**

The majority of spatial arbitrage is done by a small fraction of the buyers in the market. There are 223,169 buyers in the data. 90% of the cross-location flips completed within 7 days are conducted by less than 1% of the buyers (n=2,082.) This is consistent with Shleifer and Vishny (1997), who argued that the majority of buyers in markets do not engage in arbitrage.

## *Testing H1*

The data allow us to observe flips with perfect accuracy, but we must infer which of these flips represent spatial arbitrage. Although our system of distinguishing spatial arbitrage from other types of flips is imperfect, there is strong evidence of spatial arbitrage in the market nonetheless. For example, 46,768 (210,149) vehicles were purchased at one location and then resold at a different location within 2 (7) days by the same trader, with the average *Flipped_PriceDifference* exceeding $1,100 ($1,200.) Thus, we conclude that there is support for H1.

## *Testing H2 and H3*

### Assumption regarding the webcast channel and bidder participation

Prior to discussing our tests of H2 and H3, we first validate a key assumption underlying H3: that bidders use the webcast channel to participate in auctions in which they otherwise would not have participated. We used a panel regression model to estimate the correlation between the number of purchases a buyer made via the webcast channel (*#WebcastPurchases*) and the number of market locations (*#Locations*) from which s/he purchased. The specification is:

$\#\text{Locations}_{l,t} = \beta_0 + \beta_1\ \#\text{WebcastPurchases}_{l,t} + \beta_2\ \#\text{TotalPurchases}_{l,t} + c_l + \sum_{year=2003}^{2009} \beta_{year}\ \text{Year}_t + \varepsilon_{l,t}$, where $l$ indexes the buyer and $t$ the year. $TotalPurchases_{l,t}$ controls for each buyer's overall purchase volume, $c_l$ represents a buyer fixed effect, and $Year_t$ are dummy variables for each year, which are included to control for a possible time trend. $R^2$ is 0.75. $\beta_1$ is positive and significant ($\beta_1 = 0.02$, robust standard error = 0.002), and a one standard deviation increase in $\#WebcastPurchases_{l,t}$ is associated with a 27% increase in $\#Locations_{l,t}$. This indicates that buyers use the webcast channel to purchase from locations from which they would not have purchased physically, which provides evidence of the validity of the assumption underlying H3.

### Testing H2 and H3

We used logistic regression to examine which factors influenced the probability that a purchased vehicle $g$ would later be spatially arbitraged (*Arbitraged* = 1). This allowed us to test H2 and H3. We set *Arbitraged* = 1 if a vehicle was flipped at a different location within $\alpha$ days of the original transaction (i.e., $DaysToFlip \leq \alpha$.) In our main results, we set $\alpha = 7$, but used different values of $\alpha$ for robustness. Because estimating a series of logistic regressions for different values of $\alpha$ is similar to estimating a semiparametric hazard model, we also used a Cox proportional hazards model in which the "hazard" occurred if a purchased vehicle was later arbitraged (Cox, 1972). These results are similar to those from the logistic regression model and are not reported.

Several factors could influence the likelihood that a purchased vehicle is later arbitraged, including seller characteristics such as limited attention, bidding competition for the vehicle, vehicle characteristics, buyer characteristics, location characteristics, and time. These are each included in the specification and described below.

Seller characteristics, including limited attention: We included $SellerInattentionProxy_{h,k,t}$ in the specification to test H2. If H2 is supported, then $SellerInattentionProxy_{h,k,t}$ will be positively correlated with the probability of a vehicle being spatially arbitraged. We included $TotalVehicles\_Seller_{k,t}$ in the specification to control for effects attributable to the overall volume of vehicles offered by seller *k* at time *t*. We also included $PctArbitraged\_Seller_k$, which is the number of vehicles sold by seller *k* that were flipped within 7 days divided by the number of vehicles sold by seller *k* across the entire sample, multiplied by 100. This variable controls for unmodeled seller characteristics that might make them prone to having their vehicles arbitraged and can be thought of as a seller "fixed" effect.

Bidding competition: $SecondHighBid\_Webcast_g$ and $Webcast\_Buyer_g$ provide information on the bidding competition for vehicle *g* and correspond to the model used to motivate H3. Recall that if $v_{g,m1}* > v_{(n-1)}$, then bidder $m_1$'s webcast participation in the original auction at location *i* will either: a) require bidder $m_a$ to pay $v_{g,m1}*$ instead of $v_{(n-1)}$ to win the vehicle, or b) cause bidder $m_1$ to win the vehicle. $SecondHighBid\_Webcast_g$ captures the first case, and $Webcast\_Buyer_g$ captures the second case. If H3 is supported, then both cases should be negatively associated with the probability of a vehicle being arbitraged. Thus, negative and significant coefficients for $SecondHighBid\_Webcast_g$ and $Webcast\_Buyer_g$ would provide support for H3. However, it is not clear whether the coefficient for $Webcast\_Buyer_g$ will: a) capture the theoretical effect proposed in the model, b) reflect the general propensity for buyers who purchase vehicles via the webcast channel to arbitrage vehicles, or c) both. Thus, we are less confident in interpreting the coefficient for $Webcast\_Buyer_g$ as a test of H3 than we are of the coefficient for $SecondHighBid\_Webcast_g$.

Vehicle characteristics: We included $NormalizedPrice_g$, which is $Price_g$ divided by $Valuation_g$, in the specification. This is because the more a vehicle costs relative to its valuation, the less likely an arbitrageur can turn a profit on it. We also included $Valuation_g$, although we had no a priori expectations of its coefficient. On one hand, vehicles with higher valuations might be more attractive to arbitrageurs due to potentially higher margins; on the other hand, sellers might pay more attention to choosing optimal selling locations for vehicles with higher valuations. We included $Valuation_g^2$ to allow for a curvilinear relationship. $Mileage_g$ is included to control for vehicle quality not otherwise captured in $Valuation_g$. We also included $Mileage_g^2$. We scaled $Valuation_g$ and $Mileage_g$ by dividing by 10,000 so that all variables were of similar magnitude. $Source\_Type(v)_g$, with v = {2,3,4,5}, are four dummy variables derived from the *SourceType* variable. We used "factory" as the base case and included dummy variables for the other four source types.

Buyer characteristics: $PctArbitraged\_Buyer_l$ is the number of vehicles flipped within 7 days by buyer *l* divided by the number of purchases made by buyer *l* across the entire sample, multiplied by 100. This controls for the propensity of the buyer to arbitrage vehicles. Its coefficient should be positive, as buyers who arbitrage more vehicles in general should be more likely to arbitrage any given vehicle.

Location characteristics: $Location(i)_g$, with *i* = {2,3,…,92,93}, are 92 dummy variables representing the location at which the transaction occurred. This controls for the possibility that spatial arbitrage is more likely to originate at certain locations than others, perhaps due to the types of vehicles sold, location-specific imbalances between supply and demand, or the geographic proximity of a location to other locations.

Time and other: $DieselPrice_t$ is the average U.S. diesel price per gallon (in dollars) for the month in which the transaction occurred. We retrieved this data from the U.S. Energy Information Administration (http://www.eia.doe.gov/.) $DieselPrice_t$ affects the cost of transporting vehicles and thus may influence the likelihood of arbitrage. $Day_t$ is the day the transaction occurs, which ranges from 1 (Jan. 1, 2003) to 2,318 (May 7, 2009.) It controls for unmodeled variables that change over time that might affect the likelihood of spatial arbitrage, such as learning by traders in the market, changes in interest rates, incentive programs offered by automotive manufacturers, etc.

The logistic regression specification is shown below. *g* indexes the vehicle, *h* indexes the vehicle's model, *k* indexes the seller, *l* indexes the buyer, *i* indexes the location where the transaction was conducted, and *t* indexes the day of the transaction. Table 1 lists descriptive statistics and correlations. The correlations between variables are relatively low, suggesting that multicollinearity is unlikely to be an issue. The lone exception is the high correlation between

$DieselPrice_t$ and $Day_t$, which reflects the rise in diesel prices over the time span of the data. Dropping $DieselPrice_t$ from the model doesn't change the results.

$$\text{Probability } (\text{Arbitraged}_{g,h,k,l,i,t} = 1 \mid X_{g,h,k,l,i,t}) = \frac{e^z}{e^z + 1}, \text{ where } z = \beta_0 + \beta_1 \text{ SellerInattentionProxy}_{h,k,t} + \beta_2 \text{ TotalVehicles\_Seller}_{k,t} + \beta_3 \text{ SecondHighBid\_Webcast}_g + \beta_4 \text{ Webcast\_Buyer}_g + \beta_5 \text{ NormalizedPrice}_g + \beta_6 \text{ Valuation}_g + \beta_7 \text{ Valuation}^2_g + \beta_8 \text{ Mileage}_g + \beta_9 \text{ Mileage}^2_g + \beta_{10} \text{ PctArbitraged\_Buyer}_l + \beta_{11} \text{ PctArbitraged\_Seller}_k + \beta_{12} \text{ DieselPrice}_t + \beta_{13} \text{ Day}_t + \sum_{v=2}^{5} \beta_{type(v)} \text{ Source\_Type}(v)_g + \sum_{i=2}^{93} \beta_{location(i)} \text{ Location}(i)_g + \varepsilon_{g,h,k,l,i,t}$$

**Table 1: Descriptive statistics and correlations.**

| | Mean | SD | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. $\text{Arbitraged}_{g,k,l,i,t}$ | 0.01 | 0.08 | 1 | | | | | | | | | | | |
| 2. $\text{SellerInattentionProxy}_{h,k,t}$ | 0.84 | 0.08 | 0.01 | 1 | | | | | | | | | | |
| 3. $\text{TotalVehicles\_Seller}_{k,t}$[a] | 2.04 | 3.73 | -0.01 | 0.18 | 1 | | | | | | | | | |
| 4. $\text{SecondHighBid\_Webcast}_g$ | 0.05 | 0.21 | -0.01 | 0.03 | 0.08 | 1 | | | | | | | | |
| 5. $\text{Webcast\_Buyer}_g$ | 0.07 | 0.26 | -0.02 | 0.02 | 0.11 | -0.06 | 1 | | | | | | | |
| 6. $\text{NormalizedPrice}_g$ | 0.99 | 0.21 | -0.03 | 0.01 | -0.00 | 0.00 | -0.02 | 1 | | | | | | |
| 7. $\text{Valuation}_g$[b] | 1.09 | 0.75 | 0.00 | -0.01 | 0.14 | 0.10 | 0.15 | -0.02 | 1 | | | | | |
| 8. $\text{Mileage}_g$[b] | 5.70 | 4.29 | 0.01 | -0.03 | -0.28 | -0.10 | -0.15 | 0.05 | -0.56 | 1 | | | | |
| 9. $\text{PctArbitraged\_Buyer}_l$[c] | 1.07 | 4.27 | 0.39 | 0.01 | -0.03 | -0.01 | -0.05 | -0.09 | -0.03 | 0.05 | 1 | | | |
| 10. $\text{PctArbitraged\_Seller}_k$[c] | 1.06 | 1.20 | 0.09 | 0.03 | -0.07 | -0.02 | -0.07 | -0.12 | -0.14 | 0.16 | 0.14 | 1 | | |
| 11. $\text{DieselPrice}_t$ | 2.49 | 0.81 | -0.00 | 0.02 | -0.03 | 0.09 | 0.10 | -0.03 | 0.02 | 0.07 | 0.00 | 0.02 | 1 | |
| 12. $\text{Day}_t$[d] | 1.16 | 0.70 | -0.00 | 0.01 | -0.03 | 0.13 | 0.14 | -0.02 | 0.01 | 0.08 | 0.00 | 0.02 | 0.78 | 1 |

[a] Scaled by dividing by 100; [b] Scaled by dividing by 10,000; [c] Measured as percentages; e.g., 1.07 is 1.07%; [d] Scaled by dividing by 1,000.

Results: In our estimation, we used clustered standard errors (clustered by location and day) to account for the possibility that the error terms for transactions occurring at the same location on the same day are correlated. We estimated the model using standard logistic regression (results reported) as well as rare events logistic regression, commonly referred to as ReLogit. Results of the ReLogit model do not differ in any meaningful way and are not reported. Results of the logistic regression specification are shown in Table 2.

| Table 2: Results of logistic regression model for whether a purchased vehicle is arbitraged. | | |
|---|---|---|
| **Variable** | **Coef. (Std. Error)** | **Illustration of Practical Significance** |
| $\beta_1$: SellerInattention Proxy$_{h,k,t}$ | 0.246 (0.016) *** | Using the example from the Data section, the Pathfinder (SellerInattentionProxy$_{h,k,t}$ = 0.9) would be 21% more likely to be arbitraged than a Maxima (SellerInattentionProxy$_{h,k,t}$ = 0.1.) |
| $\beta_2$: TotalVehicles_Seller$_{k,t}$ | 0.002 (0.002) | No significant relationship. |
| $\beta_3$: SecondHighBid_ Webcast$_g$ | -0.060 (0.015) *** | SecondHighBid_Webcast$_g$ = 1 associated with 5.8% decrease in arbitrage probability. |
| $\beta_4$: WebCast_Buyer$_g$ | -0.740 (0.026) *** | Webcast_Buyer$_g$ = 1 associated with 52.2% decrease in arbitrage probability. |
| $\beta_5$: NormalizedPrice$_g$ | -0.662 (0.016) *** | 1 percentage point increase associated with 0.7% decrease in arbitrage probability. |
| $\beta_6$: Valuation$_g$ | 0.810 (0.026) *** | Increase from $10,000 to $20,000 associated with 77.5% increase in arbitrage probability. Inflection point in curvilinear relationship reached at approximately $25,800. |
| $\beta_7$: Valuation$_g^2$ | -0.160 (0.008) *** | |
| $\beta_8$: Mileage$_g$ | 0.114 (0.004) *** | Increase from 10,000 to 20,000 associated with 11.9% increase in arbitrage probability. Inflection point in curvilinear relationship reached at approximately 130,000. |
| $\beta_9$: Mileage$_g^2$ | -0.004 (0.000) *** | |
| $\beta_{10}$: PctArbitraged_Buyer$_l$ | 0.140 (0.000) *** | 1 percentage point increase associated 15.0% increase in arbitrage probability. |
| $\beta_{11}$: PctArbitraged_Seller$_k$ | 0.111 (0.003) *** | 1 percentage point increase associated 11.7% increase in arbitrage probability. |
| $\beta_{12}$: DieselPrice$_t$ | -0.071 (0.011) *** | $1 dollar increase in diesel price associated with 6.8% decrease in arbitrage probability. |
| $\beta_{13}$: Day$_t$ | -0.111 (0.013) *** | Increase from day 1 to day 2,318 associated with 22.8% decrease in arbitrage probability. |
| SellerType dummies | included | |
| Location dummies | included | |
| Pseudo-$R^2$ | 0.35 | |
| Log pseudolikelihood | -766,679 | |
| n = 31,805,961 | | * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ |

The coefficient for *SellerInattentionProxy$_{h,k,t}$* ($\beta_1$ = 0.246) is positive and significant, providing support for H2. The example from the Data section provides an illustration of the practical significance of $\beta_1$. The Nissan Pathfinder (*SellerInattentionProxy$_{h,k,t}$* = 0.9) in seller *k*'s inventory would be 21% more likely to be arbitraged than a Nissan Maxima (*SellerInattentionProxy$_{h,k,t}$* = 0.1.) The coefficient for *SecondHighBid_Webcast$_g$* ($\beta_3$ = -0.060) is negative and significant, providing support for H3. The magnitude of $\beta_3$ indicates that vehicles for which *SecondHighBid_Webcast$_g$* = 1 are 5.8% less likely to be arbitraged. The coefficient for *Webcast_Buyer$_g$* ($\beta_4$ = −0.740) is also negative and significant, providing further evidence for H3. However, as noted earlier, this evidence is more equivocal than the evidence provided by $\beta_2$. Coefficients for the control variables are consistent with expectations. For example, *NormalizedPrice$_g$* ($\beta_3$ = -0.662) is negative and significant, and the coefficients for *PctArbitraged_Buyer$_l$* ($\beta_8$ = 0.140) and *PctArbitraged_Seller$_k$* ($\beta_9$ = 0.111) are positive and significant.

<u>Robustness Checks:</u> To assess whether our results were sensitive to the threshold we used for delineating spatial arbitrage from other types of flips, we reran the logistic regression model for different values of *α*. Figure 3 plots the coefficients for *SellerInattentionProxy$_{h,k,t}$*, *SecondHighBid_Webcast$_g$*, and *Webcast_Buyer$_g$* at *α*=1, *α*=7, *α*=14, and

$\alpha$=21. Coefficients are shown in black and the 95% confidence intervals are shown in gray. The x-axis represents $\alpha$, and the y-axis represents the coefficient and confidence interval estimates.

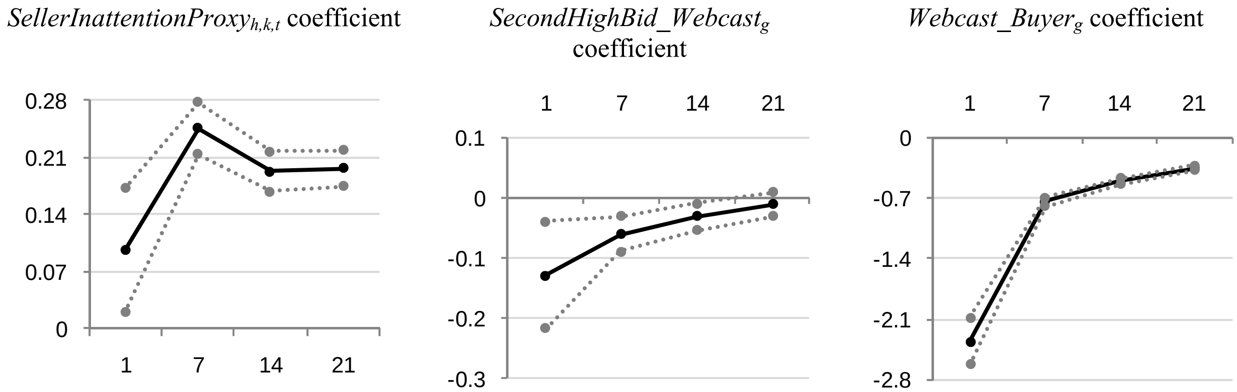


**Figure 3: *SellerInattentionProxy*$_{h,k,t}$, *SecondHighBid_Webcast*$_g$, and *Webcast_Buyer*$_g$ coefficients and 95% confidence intervals (y-axis) from logistic regression models at different values of $\alpha$ (x-axis).**

The $SellerInattentionProxy_{h,k,t}$ coefficient is positive and significant at all levels of $\alpha$. The coefficients for the variables capturing bidding competition, $SecondHighBid\_Webcast_g$ and $Webcast\_Buyer_g$, become less negative at increasing levels of $\alpha$. The $SecondHighBid\_Webcast_g$ coefficients are negative and significant at all levels of $\alpha$, except for $\alpha$ = 21, while the $Webcast\_Buyer_g$ coefficients are negative and significant at all levels of $\alpha$. We draw three conclusions from these robustness checks. First, the results are robust to different definitions of what constitutes spatial arbitrage. Second, the bidding competition effects attenuate at higher levels of $\alpha$, indicating that they only affect flips that are completed quickly. This makes sense given that bidding competition should only affect flips motivated by arbitrage and not flips motivated by other factors such as failure to sell a vehicle in the retail market, which take longer to complete. Third, the consistency of the results achieved with $\alpha$ = 1 and $\alpha$ = 7 provides evidence that the results are not confounded by the possibility that arbitrages are improving the vehicles, because it would be unlikely for an arbitrageur not only to transport but also to improve a vehicle in a single day.

### *Testing H4*

We used a discrete choice model to test H4. For every vehicle that is flipped (n=210,149), the arbitrageur must choose a destination location. We modeled the utility V of each potential location as follows.

$$V_{g,h,i,j,t} = \beta_1 * Distance_{i,j} + \beta_2 * Distance_{i,j} * SecondHighBid\_Webcast_g + \beta_3 * Distance_{i,j} * WebCast\_Buyer_g + \beta_4 * Distance_{i,j} * DieselPrice_t + \beta_5 * LaggedPrice_{h,j,t} + \beta_6 * Supply_{h,j,t} + \beta_7 * Supply^2_{h,j,t} + \beta_8 * OverallVolume_{j,t}$$

where $Distance_{i,j}$ is the distance between the source location and the potential destination location $j$, $DieselPrice_t$ is the average price per gallon of diesel fuel during the month in which the flip originated (data from http://www.eia.doe.gov/), $LaggedPrice_{h,j,t}$ is the average normalized price (i.e., price divided by valuation) of vehicles of model $h$ sold at location $j$ over the 90 days prior to time $t$, $Supply_{h,j,t}$ is the number of other vehicles of model $h$ being offered at location $j$ at time $t$, and $OverallVolume_{j,t}$ is the total number of vehicles sold at location $j$ over the 90 days prior to time $t$. Other variables are defined as above. As is standard in discrete choice modeling, we assumed that the arbitrageur chooses the location that yields the highest utility (e.g., Train 2003.) We limited the choice set to locations within 1,000 miles of the source location, as only 3.5% of arbitraged vehicles were transported more than 1,000 miles. This restriction eliminates locations from the choice set which essentially have a 0% chance of being chosen, which helps prevent the results from being contaminated by the inclusion of unrealistic potential locations and facilitates model convergence.

We interacted $SecondHighBid\_Webcast_g$, $WebCast\_Buyer_g$, and $DieselPrice_t$ with $Distance_{i,j}$ for two reasons. First, because these variables do not vary across locations, they must be interacted with a variable that does in order to enter the model. Second, we expect them to moderate the (dis)utility of distance. Because we expect arbitrageurs to prefer nearby locations over distant locations, ceteris paribus, $\beta_1$ should be negative. However, $\beta_1$ should be less negative if $SecondHighBid\_Webcast_g$ = 1 or $WebCast\_Buyer_g$ = 1, because webcast bidding activity should cause the arbitrageur to have to choose a more distant destination location (see motivation for H4.)[9] Thus, positive and significant estimates for $\beta_2$ and $\beta_3$ would provide support for H4. We expected the coefficient of $DieselPrice_t$ ($\beta_4$) to be negative, because higher fuel costs should add to the disutility of distance. We estimated the model using a multinomial logit specification. Results appear in Table 3.

**Table 3: Results of discrete choice model for arbitrageurs' choice of destination location.**

| Variable | Coef. (Std. Error) |
|---|---|
| $\beta_1$: $Distance_{i,j}$ | -0.008 (0.000) *** |
| $\beta_2$: $Distance_{i,j}$ * $SecondHighBid\_Webcast_g$ | 0.002 (0.000) *** |
| $\beta_3$: $Distance_{i,j}$ * $WebCast\_Buyer_g$ | 0.005 (0.000) *** |
| $\beta_4$: $Distance_{i,j}$ * $DieselPrice_t$ | -0.001 (0.000) *** |
| $\beta_5$: $LaggedPrice_{h,j,t}$ | 0.508 (0.015) *** |
| $\beta_6$: $Supply_{h,j,t}$ | 0.020 (0.001) *** |
| $\beta_7$: $Supply^2_{h,j,t}$ | -0.001 (0.000) *** |
| $\beta_8$: $OverallVolume_{j,t}$ | -0.000 (0.000) *** |
| | * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$ |

$\beta_2$ and $\beta_3$ are positive and significant, which provides support for H4. To illustrate the practical significance of the estimates, we calculated the elasticity of distance for the Las Vegas destination location. A 1% increase in distance from the source location to Las Vegas is associated with a 2.06% decrease in the likelihood that the arbitrageur will choose Las Vegas as the destination. However, this elasticitity decreases from -2.06% to -1.53% for $SecondHighBid\_Webcast_g$ = 1 and from -2.06% to -0.78% for $WebCast\_Buyer_g$ = 1. As expected, $\beta_4$ is negative and significant. The coefficient for $LaggedPrice_{h,j,t}$ ($\beta_5$) is positive and significant, indicating that arbitrageurs choose destinations where prices for the vehicle model *h* they are arbitraging have been relatively high. The coefficient for $Supply_{h,j,t}$ follows a curvilinear relationship, indicating that arbitrageurs prefer there to be some vehicles of the same model *h* as the vehicle they are flipping, but not too many.

## Discussion

Our research contributes to the literature on: a) the effect of electronic trading on market efficiency, and b) market integration and spatial arbitrage.

### *Electronic Trading and Market Efficiency*

Our results provide evidence that electronic trading is making the market more efficient. Because we are studying a market in which prices are determined by auction, we follow Dasgupta and Maskin (2000) and measure market efficiency based on whether each auction results in the good being sold to the bidder with the highest valuation. If the good is sold to anyone else, this is considered an inefficient outcome. The fact that spatial arbitrage occurs provides evidence of inefficiency in the wholesale automotive market, because for spatial arbitrage to occur, there must be a bidder in the market who values the vehicle more than any of the bidders present when the vehicle is initially auctioned. This inefficiency is likely due to transaction costs that prevent bidders from participating in auctions at remote locations. Electronic trading reduces these transaction costs, improving the likelihood that the

[9] $WebCast\_Buyer_g$ should be interpreted differently in the discrete choice model than in the logistic regression model, because the former uses only arbitrage transactions while the latter uses all transactions. Thus, in the discrete choice model, $WebCast\_Buyer_g$ represents whether the *arbitrageur* purchased the vehicle via the webcast channel. In the logistic regression model $WebCast\_Buyer_g$ represents whether the buyer, arbitrageur or otherwise, purchased the vehicle via the webcast channel.

bidder with the highest valuation will win the initial auction. This should improve market efficiency and erode spatial arbitrage opportunities. Our results indicate that this is happening.

By focusing on spatial arbitrage and conducting a transaction-level analysis, our results extend the existing research on electronic trading and market efficiency. Much of the existing research has used price dispersion as a proxy for market efficiency and conducted analysis using aggregate data. Our transaction-level analysis allows us to observe the underlying mechanism by which electronic trading affects market efficiency. Observing the behavioral mechanism is important because it shows how micro-level behavior leads to macro-level effects.

The overall volume of electronic transactions is growing within the market. This suggests that the total amount of spatial arbitrage should be declining. To examine this, we counted the spatial arbitrage transactions for each year from 2003 to 2008. (We dropped 2009 because we do not have data for the entire year.) Figure 4 shows the number of spatial arbitrage transactions by year, which is declining, and the percentage of electronic transactions, which is rising.

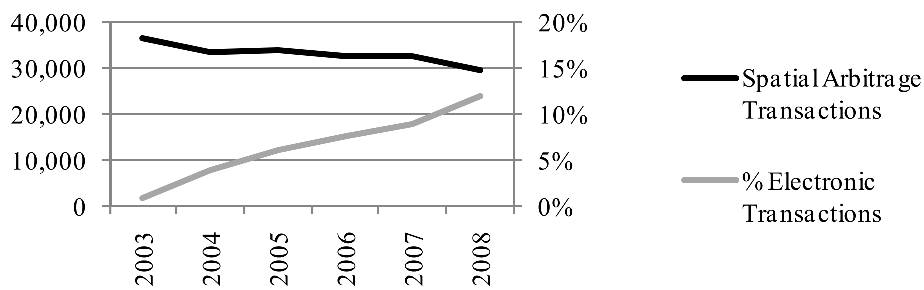


**Figure 4: Arbitrage and electronic transaction summary statistics by year.**

This result is consistent with the interpretation that arbitrageurs act as "market-makers" who provide liquidity to the market by taking positions on assets and matching buyers and sellers. Our results indicate that this function of the arbitrageur is becoming less valuable as electronic trading becomes more prevalent. This is consistent with findings reported in Hendershott and Moulton (2007), who found that floor broker and specialist trading volumes on the NYSE declined significantly after the introduction of the OpenBook system, which electronically delivers limit order book information to traders off of the physical exchange floor.

### *Market Integration and Spatial Arbitrage*

We extend the literature on market integration and spatial arbitrage in two ways. First, existing studies in this stream have generally relied on price data to infer the existence and prevalence of spatial arbitrage, despite the fact that the flow of products between locations is a critical aspect of the phenomenon (Barrett 2005.) The transaction-level granularity of our data allows us to overcome this limitation and improves the precision of our analysis. We find that speculation occurs on approximately 0.7% of transactions. This represents a lower bound on the true degree of speculation, as we do not observe speculative activity that either originates or terminates at auction companies other than the one that provided the data. Second, existing studies in this stream do not examine the behavioral factors that generate the price disparities across locations. This is because the coarseness of the data typically used prevents direct observation of the arbitrageurs' and sellers' behavior. We improve upon this by observing arbitrage activity at the transaction-level and by hypothesizing and finding evidence that sellers' limited attention is partly responsible for the price disparities that the arbitrageurs are exploiting.

## Conclusion

We used an extensive transaction-level data set to examine spatial arbitrage and the effect of electronic trading in the context of the wholesale automotive market. Results indicate that sellers' bounded rationality due to limited attention is one of the factors that contributes to the spatial arbitrage opportunities exploited by arbitrageurs. Electronic trading has improved market efficiency by allowing any buyer to source a vehicle from a low-priced location and move it to a high-priced location, a role which was previously reserved for arbitrageurs. This has made spatial arbitrage more difficult.

Although we find that electronic trading is associated with less spatial arbitrage, this may not always be the case. It is possible that, in some contexts, electronic trading facilitates spatial arbitrage. For example, spatial arbitrageurs might purchase products in bulk from one location and then use an electronic channel such as eBay to resell them at a profit to buyers in other locations. Examining spatial arbitrage in such a context and determining the contextual factors that influence whether electronic trading has a positive or negative effect on spatial arbitrage is an opportunity for future research.

Although our analysis is specific to the wholesale automotive market, the results can be generalized to markets for other products such as agricultural commodities, building materials, industrial equipment and machinery, and metals. In each of these markets, supply/demand conditions and bounded rationality due to limited attention are likely to create pricing disparities across locations for arbitrageurs to exploit. The automotive market is well-suited for examining these issues because each product is uniquely identifiable via its VIN, thereby permitting us to observe arbitrage at the transaction level. Similar identifiers are available in other markets and will become increasingly available as identification and tracking technologies such as RFID become more widely adopted. Future research might use methods similar to ours to estimate the incidence of spatial arbitrage in other markets and whether spatial arbitrage is affected by electronic trading.

## References


Bakos, J. Y. 1998. "The Emerging Role of Electronic Marketplaces on the Internet," *Communications of the ACM* (41:8), August, pp. 35-42.

Baulch, B. 1997. "Transfer Costs, Spatial Arbitrage, and Testing For Food Market Integration," *American Journal of Agricultural Economics* (79:2) May, pp. 477-487.

Baye, M., Morgan, J., and Scholten, P. 2006. "Information, Search, and Price Dispersion," in *Handbook on Economics and Information Systems*, T. Henderschott (ed.), Amsterdam: Elsevier.

Barrett, C. B. 2005. "Spatial Market Integration," *The New Palgrave Dictionary of Economics*, 2nd Edition. London: Palgrave Macmillan.

Brynjolfsson, E. and M. D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science* (46:4), April, pp. 563-585.

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), May, pp. 332-350.

Corwin, S., and Coughenour, J. 2008. "Limited Attention and the Allocation of Effort in Securities Trading," *Journal of Finance* (63:6), December, pp. 3031-3067.

Cox, D. R. 1972. "Regression Models and Life Tables," *Journal of the Royal Statistical Society, Series B* (34:2), pp. 187-202.

Dasgupta, P. and Maskin, E. 2000. "Efficient Auctions," *The Quarterly Journal of Economics* (115:2), May, pp. 341-388.

Fackler, P. and Goodwin, B. 2001. "Spatial Price Analysis," in *Handbook of Agricultural Economics, vol. 1B*, B. Gardner and G. Rausser (eds.), Amsterdam: Elsevier.

Hendershott, T. and Moulton, P.C. 2007. "Speed and Stock Market Quality: The NYSEs Hybrid," Working paper, University of California, Berkeley.

Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Lee, H. G. 1998. "Do Electronic Marketplaces Lower the Price of Goods?" *Communications of the ACM* (41:1), January, pp. 73-80.

Lee, H. G., Westland, J. C., and Hong, S. 1999. "The Impact of Electronic Marketplaces on Product Prices: An Empirical Study of Aucnet," *International Journal of Electronic Commerce* (4:2), pp. 45-60.

Malone, T. W., Yates, J., and Benjamin, R. I. 1987. "Electronic Markets and Electronic Hierarchies," *Communications of the ACM* (30:6), June, pp. 484-497.

Milgrom, P. R. and Weber, R. J. 1982. "A Theory of Auctions and Competitive Bidding," *Econometrica* (50:5), September, pp. 1089-1122.

Peng, L. and Xiong, W. 2006. "Limited Attention and Asset Prices," *Journal of Financial Economics* (80:3), June, pp. 563-602.

Schuler, R. E. 2008. "Spatial Arbitrage To Reduce Seams Across Electricity System Control Areas: An Experimental Analysis," Working paper, University of Trier.

Sharpe, W., and Alexander, G. 1990. *Investments*, 4th Edition, Englewood Cliffs, N.J.: Prentice Hall.

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

Simon, H. 1955. "A Behavioral Model of Rational Choice," *The Quarterly Journal of Economics* (69:1), February, pp. 99-118.

Train, K. E. 2003. *Discrete Choice Methods with Simulation*. Cambridge: University Press.