MANAGEMENT SCIENCE
Vol. 58, No. 2, February 2012, pp. 394–412
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)
http://dx.doi.org/10.1287/mnsc.1110.1494

A Transaction-Level Analysis of Spatial Arbitrage: The Role of Habit, Attention, and Electronic Trading

Eric Overby, Jonathan Clarke
College of Management, Georgia Institute of Technology, Atlanta, Georgia 30308
{eric.overby@mgt.gatech.edu, jonathan.clarke@mgt.gatech.edu}

Despite the central role of arbitrage in finance and economic theory, there is limited evidence of the factors that create and eliminate arbitrage opportunities, how often arbitrage occurs, and how profitable it is. We address these gaps via a transaction-level analysis of spatial arbitrage in the wholesale automotive market. We investigate why arbitrage opportunities are created by analyzing how sellers choose where to sell vehicles. We find that the attention sellers pay to the distribution of a vehicle is negatively related to the probability that it is arbitraged. Arbitrage occurs in approximately 1% of transactions, although electronic trading is making arbitrage less prevalent by improving buyer/seller matching across locations. Arbitrage yields a 5.6% return on average, although arbitrageurs take a loss 14% of the time. Our results contribute to the literature on arbitrage, the effect of attention allocation on market outcomes, and the effect of information technology on market efficiency.

Key words: spatial arbitrage; seller distribution; decision making; habit; attention allocation; market efficiency; electronic trading; information technology; automotive
History: Received July 15, 2010; accepted October 23, 2011, by Brad Barber, Teck Ho, and Terrance Odean, special issue editors. Published online in *Articles in Advance* January 13, 2012.

1. Introduction

Arbitrage is one of the most fundamental concepts in finance and economics. If prices fail to reflect the fundamental value of assets, then theory suggests that arbitrageurs will exploit these mispricings, quickly returning prices to fundamental values. Arbitrage is central to many theories, including the efficient markets hypothesis and the "law of one price," although research in behavioral finance has argued that the role of arbitrage in keeping markets efficient is more limited than traditional theory would suggest (e.g., Shleifer and Vishny 1997). Despite arbitrage's central place within theory, data coarseness has made it difficult to observe. Thus, there is limited empirical evidence of the factors that create and eliminate arbitrage opportunities, how often arbitrage occurs, and how profitable it is. We address this gap by conducting a transaction-level examination of a specific type of arbitrage: spatial arbitrage. Following Coleman (2009), we define "spatial arbitrage" as the purchase and subsequent resale of the same product in different geographic locations in order to exploit a price discrepancy. This mirrors the textbook definition of "arbitrage," which is "the simultaneous purchase and sale of the same, or essentially similar, securities in two different markets for advantageously different prices" (Sharpe et al. 1995, p. 1001), except that spatial arbitrage does not involve simultaneous transactions given the need to transport the product. Despite this distinction, spatial arbitrage plays the same theoretical role as arbitrage with respect to market efficiency. To wit, if prices for the same product differ across locations by more than the cost of transportation between the locations, then spatial arbitrage opportunities will exist until prices reconverge (Fackler and Goodwin 2001).

We study spatial arbitrage in the context of the United States wholesale automotive market, which is well suited for our study for three reasons. First, market locations are distributed throughout the United States, and average prices for vehicles of the same year, make, and model often differ across locations by more than the cost of transportation, creating opportunities to spatially arbitrage these vehicles. Second, we can observe spatial arbitrage in this market with a high degree of precision. This is because the traders, locations, and products in the market are each uniquely identified, with the vehicle identification number ("VIN") serving as the unique identifier for each product. This allows us to observe traders who engage in spatial arbitrage by buying a vehicle at location i for price p and then reselling the same vehicle at location j for a higher price p'. Third, the industry has steadily been transitioning from physical trading to electronic trading. Our data span a time period (January 2003–May 2009) during which electronic trading increased from approximately 0% to 14%. This allows us to examine whether information technology is having an effect on spatial arbitrage by reducing traditional market frictions.

We pose the following research questions. First, why do spatial arbitrage opportunities arise? Second, does information technology influence spatial arbitrage by reducing market frictions? Third, how common is spatial arbitrage, how large and variable are the profits from it, and what percentage of traders engage in it?

Our results contribute to three distinct research streams. First, we contribute to the growing literature about how the allocation of decision makers' attention affects market outcomes (e.g., Barber and Odean 2008, Corwin and Coughenour 2008, Peng and Xiong 2006) by analyzing why spatial arbitrage opportunities arise. We do this by examining how sellers in the market choose the selling location of vehicles. Consistent with research on how habit affects decision making (e.g., Samuelson and Zeckhauser 1988), we find that habit plays a large role in how sellers distribute vehicles across locations: sellers tend to offer vehicles of a given year/model at locations where they have offered them in the past. This force of habit is likely to result in vehicles sometimes being offered at locations where they fetch prices below those available at other locations, thereby creating opportunities for spatial arbitrage. However, the influence of habit is moderated by the amount of attention sellers allocate to vehicles of each year/model (e.g., 2007 Chevrolet Malibus, 2003 Honda Accords) in their inventory, which we measure as the prevalence of the vehicle year/model in a seller's inventory.[1] This indicates that sellers are more likely to break with habit for vehicles that are prevalent in their inventory, which we interpret as evidence of more careful distribution of these vehicles. Consistent with this, we find that these vehicles are less likely to be arbitraged.

Second, we contribute to the literature on how information technology affects market efficiency by reducing market frictions (e.g., Brown and Goolsbee 2002, Jensen 2007). Existing research in this stream has typically used price dispersion as the measure of efficiency, at least in part because more granular data is not available. By contrast, we use the prevalence of spatial arbitrage to measure efficiency. Spatial arbitrage is a more microlevel measure because it reflects the behavior of the individual traders most aware of potential inefficiencies, whereas price dispersion is a more macrolevel measure of overall market conditions that may contain noise due to unobserved transaction costs or quality differences.[2] Our data span more than six years from January 2003 to May 2009. During this time, the percentage of transactions conducted electronically rose from approximately 0% to 14%, and the percentage of arbitrage transactions declined by 13%. A key benefit of electronic trading is that it gives buyers the reach to purchase directly from locations where prices are low (e.g., Overby and Forman 2011), rather than relying on arbitrageurs to purchase from these locations and move the vehicles to them. In this way, electronic trading allows buyers to disintermediate the arbitrageurs. We find that electronic trading is associated with reduced spatial arbitrage (and thereby greater efficiency) and also makes spatial arbitrage more difficult by requiring arbitrageurs to move vehicles greater distances.

Third, we contribute to the empirical arbitrage literature by leveraging our transaction-level data to develop estimates of the extent of arbitrage activity relative to overall trading and the magnitude and variance of arbitrage profits. These estimates have been elusive in prior research because of the coarseness of data used. The granularity of our data allows us to estimate these values as well as to test two of the claims about arbitrage offered by Shleifer and Vishny (1997): (a) that relatively few traders engage in arbitrage, and (b) that arbitrage requires capital and is risky. Our results indicate that approximately 1% of sold vehicles are arbitraged. Relatively few traders engage in arbitrage; 1% of the traders account for over 90% of the arbitrage transactions. According to our estimates, arbitrageurs incur costs of $10,700 per arbitraged vehicle and net a profit of $600 (on average), which represents an average accounting return of 5.6%. The ten most active arbitrageurs in the data average 496 arbitrage transactions per year each (together, they account for 15% of all arbitrage activity); these arbitrageurs achieve an estimated annual profit of $311,000 each. However, arbitrage carries risk, as profits are variable (the standard deviation is approximately $800) and arbitrageurs take a loss approximately 14% of the time. The seller of a vehicle that is later arbitraged forfeits approximately $1,100 in forgone revenue. This is a trivial amount for any individual seller compared to her overall revenue, but forms a substantial industry for arbitrageurs when aggregated across all sellers.

This paper proceeds as follows. In §2, we present a brief review of the empirical research on spatial arbitrage and arbitrage in financial markets. In §3, we describe the empirical context, present a summary of the data, and discuss how we identify spatial arbitrage transactions. In §§4 and 5, we examine why spatial arbitrage opportunities arise and the effect that the steady increase in electronic trading in the market

[1] We also measure attention based on how much prices for vehicles of a year/model vary across locations. We found this measure to be insignificant.

[2] Some of the price dispersion papers use the term arbitrage. For example, Jensen (2007) referred to sellers' decisions to shift supply between locations in order to exploit price discrepancies as arbitrage. We submit that this behavior is better described as optimal distribution than as arbitrage, because arbitrage—as we define it—requires both a purchase and a sale.

is having on arbitrage opportunities. In §6, we present our estimates of the magnitude and variance of arbitrage profits. Section 7 provides a summary of the results and contributions along with limitations and opportunities for future research.

2. Brief Review of Empirical Research on Arbitrage

To our knowledge, this is the first transaction-level study of spatial arbitrage and of arbitrage in a more general sense. Traditionally, research on spatial arbitrage has estimated the prevalence of spatial arbitrage by analyzing the comovement of time series of prices at different locations (e.g., Alexander and Wyeth 1994, Ravallion 1986). This method is based on the idea that if prices do not move together, then the resulting cross-location price differences will create opportunities for spatial arbitrage. Researchers have referred to this as a test of the degree to which a market is "integrated." More recent research has expanded this approach by examining "spatial market integration" using not only prices but also the transaction costs associated with trading across locations (e.g., Barrett 2008, Baulch 1997). Because of the coarseness of data used in this stream, none of these studies are based on observable arbitrage transactions by individual traders. Financial market studies of arbitrage have similarly been limited by data coarseness. One example is Pontiff (1996), who inferred arbitrage activity from the price discounts on closed-end funds rather than observation of individual trades. Also, studies of the impact of market frictions on arbitrage activity (Mitchell et al. 2007, Mitchell and Pulvino 2001) are based on price data and infer arbitrage activity from hedge fund returns. Although these studies provide estimates of the returns to various arbitrage strategies, they are not based on observation of individual arbitrage trades. The detail available in our transaction-level data (including unique identifiers for traders, products, and locations) permits us to observe instances in which a specific trader engages in spatial arbitrage by purchasing an item at location i at time t and then quickly reselling that same item at location j at time t'. This granularity allows us to explore many largely unanswered questions about arbitrage, including why arbitrage opportunities occur, how electronic trading influences arbitrage, the extent of arbitrage activity relative to overall trading, and the magnitude and variance of arbitrage profits.

3. Empirical Context, Data Summary, and Delineation of Spatial Arbitrage

3.1. Empirical Context

We use the United States wholesale automotive market as the context in which to study spatial arbitrage.[3] This market is a business-to-business market for the exchange of used vehicles. 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.[4] Dealers range in size from small independent dealerships to large national chains such as CarMax. 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. We refer to these dealers as "arbitrageurs," and we refer to the purchase and resale transactions they conduct as "arbitrage flips." Sellers in the market are one of two types (broadly speaking): institutional sellers and dealer sellers. First, institutional sellers include automotive manufacturers (e.g., Toyota, Ford) and their finance arms (e.g., Toyota Financial Services, Ford Credit), rental car companies (e.g., Hertz), and banks (e.g., Bank of America). These firms choose to sell in the wholesale market because many lack retail outlets and thus are not equipped to sell to retail consumers. Also, the wholesale market is generally more liquid and predictable than the retail market, which allows these firms to dispose of multiple vehicles quickly at a predictable price. Second, dealer sellers are used car dealers who use the market to sell vehicles. Perhaps the most common reason for a dealer to be a seller in the wholesale market is if he cannot 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 he is an arbitrageur completing the sale end of an arbitrage flip. (Hereafter, we use the term "seller" to refer to sellers who are not engaged in arbitrage.)

Approximately nine million vehicles are traded in the wholesale market each year (National Auto Auction Association 2011). Trading has traditionally occurred 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 these 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 from the buyers in an open outcry ascending auction. The seller has the option to accept or reject the high bid. Most transactions at each facility are conducted on

[3] We use the term "market" in a holistic sense to refer to the overall system used for the wholesale trade of vehicles in the United States. We use the term "location" to denote the city in which trades occur.

[4] Used car dealers obtain about 50% of the vehicles they sell as trade-ins and the other 15% in miscellaneous ways (National Automobile Dealers Association 2009).

the same day each week, referred to as "sale day." Average prices for vehicles of the same year, make, and model often vary across facilities because of supply/demand imbalances and other market frictions.

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 log into the webcast from an Internet browser to receive live audio/video of the physical auction and to place bids in competition with buyers who are physically present at the facility. This gives buyers the option to participate in the auctions and to place bids electronically if they cannot (or choose not to) travel to the physical auction facility. It is worth noting that the webcast channel augments the physical auction process, and thus is not a pure electronic channel. Also, the webcast channel does not affect the mechanism used to determine prices. A human auctioneer solicits bids until the highest bid is registered, regardless of whether the bids are placed by buyers using the physical or webcast channels. Also, the webcast channel does not fundamentally change sellers' behavior. Sellers transport vehicles to market facilities to be auctioned, and they either accept or reject the high bid for each vehicle regardless of the channel by which it was placed.[5]

3.2. Summary of the Data

Data were provided by a major automotive auction company and consist of all vehicles auctioned ($n = 53{,}691{,}888$) at the company's physical facilities between January 1, 2003, and May 7, 2009. In all, 31,805,961 (59.2%) vehicles were sold. Of these, 7.5% were sold to buyers using the webcast channel, with this percentage growing from below 1% in 2003 to 14.5% in 2009. For each vehicle auctioned, the data contain an indicator of whether the vehicle was sold (*Sold*?), an identification number for the seller (*SellerID*), the vehicle identification number (*VIN*), the auction date (*Date*), the vehicle's make and model (*VehicleModel*), the vehicle's model year (*VehicleYear*), the vehicle's odometer reading (*Mileage*), the location at which the vehicle was auctioned (*LocationID*) along with the location's zip code, and an indication of the type of seller offering the vehicle (*SellerType*). There are 93 locations in the data, which are distributed throughout the United States. *SellerType* takes one of five values, the first four of which represent institutional sellers. These values are (1) factory, which refers to automotive manufacturers that sell vehicles previously used as company cars; (2) lease, which refers to leasing and finance companies that sell vehicles whose lease has expired; (3) rental, which refers to rental car firms that sell vehicles retired from rental service; (4) repossession, which refers to financial institutions that sell vehicles that have been repossessed; and (5) dealer, which refers to dealer sellers. The percentage of vehicles offered by *SellerType* are 12.5% (factory), 22.6% (lease), 3.8% (rental), 5.3% (repossession), and 55.8% (dealer). There are 179,957 unique *SellerID*'s in the data. The average number of vehicles auctioned per seller is 299, with a standard deviation of 4,594 and a median of 15. This high skewness reflects the diversity of sellers in terms of their size (e.g., Hertz sells more vehicles in the market than does a small, independent car dealer).

For each vehicle sold, the data include an identification number for the buyer (*BuyerID*), the buyer's zip code (*BuyerZip*), the transaction fees paid by the buyer and seller (*BuyFee* and *SellFee*), the transaction price (*Price*), and the vehicle's estimated wholesale valuation (*Valuation*). *Valuation* is calculated by the auction company based on transactions for similar vehicles over the prior 30 days. The data also contain the channel (physical or webcast) through which the buyer purchased the vehicle. We use a dummy variable (*Webcast_Buyer*) to denote whether a vehicle was purchased by a buyer using the webcast channel. 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 (*Webcast_SecondHighestBid*). The auction company records this data to help quantify the impact of the 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 $Webcast\_SecondHighestBid = 0$. There are 223,969 unique *BuyerID*'s in the data. The average number of vehicles purchased per buyer is 142, with a standard deviation of 495 and a median of 14. See the appendix for a list of variable definitions.

[5] There are also pure electronic channels that operate within the industry. These channels operate similarly to eBay; sellers post vehicle listings on webpages, and buyers place bids on vehicles within a specified time window or purchase the vehicle at a posted price. We exclude transactions conducted in these channels for three reasons. First, they are rare; less than 1% of the vehicles in our sample are purchased via these channels. Second, the posted price option and the lack of a human auctioneer in these channels alters the price discovery mechanism; including pure electronic transactions would introduce this confounding factor into our analysis. Third, sellers do not have to choose the facility at which to sell vehicles when using the pure electronic channels. As discussed in §4, modeling this choice is a key part of our analysis.

3.3. Delineation of Spatial Arbitrage

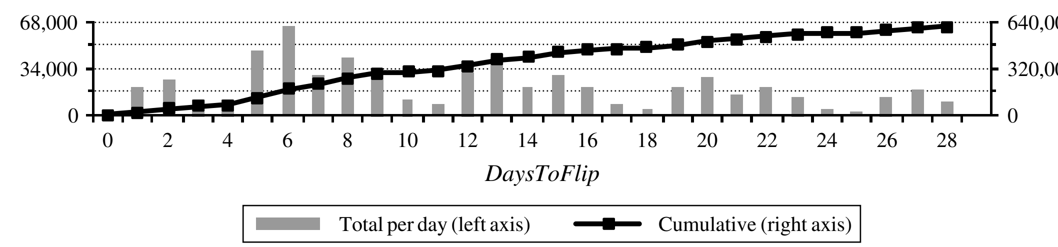
We define a "flip" as a pair of transactions for the same vehicle (as identified by *VIN*) in which the buyer in the first transaction is the seller in the second transaction (as identified by *BuyerID* and *SellerID*). A buyer may flip a vehicle for several reasons. First, a buyer may flip a vehicle because he is engaging in spatial arbitrage. Second, a buyer 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 buyer may flip a vehicle if he is unable to sell it in the retail market and chooses to liquidate it via the wholesale market. Fourth, a buyer may flip a vehicle if 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. There are 2,123,718 flips in the data.

We delineate arbitrage flips from other types of flips as follows. We focus on "cross-location flips," or flips in which the vehicle is moved from the location at which it was purchased (i.e., the "source" location) to a different location for resale (i.e., the "destination location"). Not only is this appropriate because of our interest in spatial arbitrage, but it also helps us differentiate arbitrage flips. For example, we cannot tell from our data if the buyer performing the flip improved the vehicle. However, we assume that if he did, he would resell the vehicle at the same location from which 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 arbitrage are actually motivated by other factors. We also used *DaysToFlip*, which is the number of days between the two transactions comprising a flip, to delineate arbitrage flips from other types of flips. We reasoned that arbitrage flips would be completed the most quickly, i.e., they would have the lowest *DaysToFlip*. This is because the goal of the arbitrageur is to maximize his profits, and each day the arbitrageur retains ownership of a vehicle reduces his profits because of his cost of capital and depreciation. Flips due to improving the vehicle would be completed the second-most quickly. This is because the dealer will still wish to minimize the number of days he retains the vehicle, but 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 dealer 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.

We classified a cross-location flip as an arbitrage flip if it was completed within α 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 α is set too high and false negatives (e.g., failing to categorize a flip as arbitrage when it should be) if α is set too low. In our main analysis, we set $\alpha = 7$, and we varied this threshold to test for sensitivity. We used a conservative (i.e., low) α threshold for our main results to limit the possibility that cross-location flips were motivated by a factor other than arbitrage. There are 210,170 arbitrage flips within the data for $\alpha = 7$, which represents 0.7% of the total transactions. Figure 1 shows the number of cross-location flips for different *DaysToFlip*. We note that our estimates of the number of arbitrage flips may be conservative because we do not observe arbitrage activity that either originates or terminates at auction companies other than the one that provided the data.

There are a total of 11,116 buyers in the data (which is 5.0% of all buyers) who conducted at least one arbitrage flip over the six-plus-year time period of the

Figure 1 **Number of Cross-Location Flips for Different *DaysToFlip***



Notes. A "cross-location flip" is a pair of transactions for the same vehicle in which the buyer in the first transaction is the seller in the second transaction and the two transactions occur at different locations. *DaysToFlip* is defined as the number of days between the two transactions that comprise a cross-location flip. The chart shows the number of cross-location flips per *DaysToFlip* (gray bars; measured on the left axis) and the total cross-location flips up to and including *DaysToFlip* (black line; measured on the right axis). We classified a cross-location flip as an arbitrage flip if it was completed within α days of the original purchase, i.e., if *DaysToFlip* $\leq \alpha$.

study. The average number of flips for these buyers is approximately 19, with a standard deviation of 136 and a median of 2. To explore this skewness further, we calculated that 50% of the arbitrage flips are conducted by 1.2% of these buyers (which represents 0.06% of all buyers), and 80% of the arbitrage flips are conducted by 7.6% of these buyers (which represents 0.38% of all buyers).

4. Why Are Spatial Arbitrage Opportunities Available?

Spatial arbitrage opportunities exist when a vehicle is sold at location i for a price lower than that available at location j, after accounting for transportation and related transaction costs. In some sense, this suggests that the vehicle's seller "misdistributed" the vehicle by selling it at location i rather than at location j. We used a discrete choice model to analyze how sellers choose how to distribute their vehicles, including how these decisions may create opportunities for spatial arbitrage.

4.1. Specification of Seller Distribution Choice Model

There are 53,691,888 vehicles auctioned in the data, each of which represents a choice by a seller to auction a vehicle at a location.[6] Sellers choose the location from a set of alternative locations. Although sellers can technically choose from among all locations in the market, none of the sellers in the data use all locations. Thus, we assume that each seller has a set of locations at which she offers vehicles; we recover this set for each seller from the data. For example, assume a seller offers a 2006 Ford Focus in Dallas on May 5, 2008, and that we observe the seller to use the Dallas, Houston, Nashville, and New Orleans locations. We assume the seller chose the Dallas location from the four alternatives in her choice set: Dallas, Houston, Nashville, and New Orleans. As is standard in choice models, we assume the seller chooses to sell each vehicle at the location from her choice set that provides the highest utility. We model utility of each location as $U_{ghikt} = V_{ghikt} + \varepsilon_{ghikt}$, where g is the vehicle, h is the vehicle's year/model, i is the location, k is the seller, t is the week, and $V_{ghikt} = \beta_1 \times Propensity_{hikt} + \beta_2 \times Propensity_{hikt} \times Attn\_Prevalence_{hkt} + \beta_3 \times Propensity_{hikt} \times Attn\_GeoPriceStDev_{ht} + \beta_4 \times Supply_{hit} + \beta_5 \times Supply^2_{hit} + \beta_6 \times PctSold_{hit} + \beta_7 \times Price_{hit} + \sum_{i=2}^{93} \beta_{8i} \times Location(i)$. We discuss each of the variables in the utility function below.

[6] Our model is conditional on the seller choosing to auction a vehicle. We do not model the seller's decision of whether to auction the vehicle or keep it in service within her fleet (Schiraldi 2011).

4.1.1. Variables Influencing the Utility of Each Location. Sellers decide the selling locations of the vehicles in their fleet on a recurring basis, with many sellers making these decisions each week. When choosing the locations at which to auction their vehicles, sellers have the option to collect and analyze information about each potential location. Both data collection and analysis may be costly for the seller. The former requires that sellers gather data on recent supply/demand patterns at each location for all vehicles in their fleet, and the latter requires that sellers use these (and potentially other) data to develop a forecast of the optimal selling locations for their vehicles. Given a seller's potential costs of making optimal distribution decisions for all her vehicles on a recurring basis, it seems reasonable that she might resort to shortcuts in her decision making. In particular, research has shown that habit plays a prominent role in recurring decisions, such that decision-makers' choices may be determined as much or more by their choices in the past as by an explicit calculation of the utility of current alternatives (e.g., Aarts et al. 1998). This behavior may be rational, depending on the cost of evaluating and switching to other alternatives (Klemperer 1987). To account for the possibility that sellers choose the selling location of vehicles based on where they have sold vehicles of the same year/model in the past, we include $Propensity_{hikt}$ as part of the utility of each location i in a seller's choice set. $Propensity_{hikt}$ is the number of vehicles of year/model h that seller k offered at location i divided by the number of vehicles of year/model h that seller k offered at all locations, both in the 52 weeks (or fewer for observations in year 2003) prior to week t.

Part of the reason that habit plays an important role in recurring decisions is that decision makers are boundedly rational and have limited attention to allocate across decisions (e.g., Kahneman 1973, Simon 1955).[7] In our context, sellers must allocate their attention to the distribution of attention to vehicles of year/model h, then the predominant factor that determines the distribution of these vehicles is likely to be habit (e.g., Samuelson and Zeckhauser 1988). Conversely, if a seller allocates a large amount of attention to vehicles of year/model h, then we expect that she will conduct a more deliberate evaluation of where to distribute vehicles, thereby relying less on habit to make her decision. We capture this by including interaction terms between $Propensity_{hikt}$ and two measures of how much attention the seller pays

[7] If decision makers are boundedly rational, then they have a finite amount of cognitive capacity to apply to decision making. We consider "attention" to be determined by how they allocate their cognitive capacity; e.g., if a decision maker allocates more of her cognitive capacity to decision A than to decision B, then we consider her to pay more attention to decision A.

Table 1 Descriptive Statistics for Variables in the Seller Distribution Choice Model

			Correlations					
	Mean	Std. dev.	1	2	3	4	5	6
1. $Propensity_{hikt}$	0.11	0.27	1					
2. $Attn\_Prevalence_{hkt}$	0.09	0.16	0.08	1				
3. $Attn\_GeoPriceStDev_{ht}$[a]	1.88	0.89	0.01	−0.02	1			
4. $Supply_{hit}$[b]	1.24	3.30	0.12	0.15	−0.03	1		
5. $PctSold_{hit}$	0.43	0.39	0.08	0.02	−0.08	0.21	1	
6. $Price_{hit}$[c]	1.08	0.72	0.02	0.03	0.63	0.07	−0.07	1

Note. See the appendix for variable definitions.
[a]Scaled by dividing by 1,000.
[b]Scaled by dividing by 10.
[c]Scaled by dividing by 10,000.

to the vehicle she is distributing. The first measure is the prevalence of the vehicle in seller k's fleet at week t ($Attn\_Prevalence_{hkt}$), as measured by the percentage of vehicles of the same year/model h offered by seller k at week t relative to all vehicles offered by seller k at week t. For example, assume seller k offered 10 vehicles in week t, 9 of which were 2007 Nissan Maximas and 1 of which was a 2007 Nissan Pathfinder. $Attn\_Prevalence_{hkt}$ would be 0.9 for the 2007 Nissan Maximas and 0.1 for the 2007 Nissan Pathfinder. The second measure is the degree to which the price for a vehicle of model/year h at week t is likely to vary across geographic locations ($Attn\_GeoPriceStDev_{ht}$). We calculated this measure using two steps. First, we calculated the mean price of vehicles of year/model h at each location they were sold in the three weeks prior to week t. Second, we calculated $Attn\_GeoPriceStDev_{ht}$ as the standard deviation of the mean prices across locations, scaled by dividing by 1,000. For example, assume that 2007 Nissan Pathfinders were sold in Chicago, St. Louis, and Pittsburgh in the three weeks prior to the fifth week of 2008 and that their mean prices at these locations were \$12,000, \$13,000, and \$11,500, respectively. In this case, $Attn\_GeoPriceStDev_{ht} = 0.764$ (note the scaling).[8] It is worth noting that the standard deviation of prices across locations should affect the amount of attention paid by a seller, but that the mean of prices should not. This is because the returns to distribution are based on how much prices for vehicles of year/model h vary across locations, not on how much those vehicles are worth on average.

The intuition behind using $Attn\_Prevalence_{hkt}$ and $Attn\_GeoPriceStDev_{ht}$ to measure attention is straightforward. Assume that sellers pay a cost to collect and analyze data about potential selling locations for each year/model h. If a seller pays this cost for a year/model that is prevalent in her fleet, then she will reap the benefits of this investment across a relatively large number of vehicles. Similarly, the returns to paying attention to the distribution of vehicles of year/model h should be increasing in the degree to which prices for those vehicles vary across locations. By contrast, there is relatively little incentive for a seller to pay attention to year/models that are rare in her fleet or that fetch similar prices regardless of where they are offered.

The other variables in the seller's utility specification are $Supply_{hit}$, $Supply^2_{hit}$, $PctSold_{hit}$, $Price_{hit}$, and $\sum_{i=2}^{93} Location(i)$. $Supply_{hit}$ is the number of vehicles of year/model h offered at location i in the three weeks prior to week t, scaled by dividing by 10. It captures the degree to which sellers consider the historical supply of similar vehicles at a location when making distribution decisions. We include $Supply^2_{hit}$ to allow for a curvilinear relationship. $PctSold_{hit}$ is the percentage of vehicles of year/model h offered at location i in the three weeks prior to week t that were sold. It captures the utility (disutility) of high (low) sales percentages at a location. $Price_{hit}$ is the average price of vehicles of year/model h sold at location i over the three weeks prior to week t, scaled by dividing by 10,000. This accounts for the possibility that sellers choose locations based on historical prices at those locations. $Price_{hit}$ is null when no vehicles of year/model h are sold at location i during the lagged period; this occurs approximately 40% of the time. We replaced these nulls with the mean of $Price_{hit}$ for all locations at which vehicles of year/model h were sold during the lagged period. $Location(i)$ are 92 dummy variables that capture the average unobserved utility of each location (referred to alternative-specific constants in the choice model literature; e.g., Train 2009). Table 1 shows descriptive statistics for the variables in the seller distribution choice model.

Another variable that affects the seller's utility is the distance between a vehicle g's location prior to entering the market and each location i ($Distance_{gi}$). This

[8] $Attn\_GeoPriceStDev_{ht}$ is null when vehicles of year/model h were sold at one or fewer locations over the lagged period, which occurred 3.5% of the time. In these cases, we calculated $Attn\_GeoPriceStDev_{ht}$ by extending the lagged period until it covered two locations where vehicles of year/model h were sold. For cases in which this did not work (0.9%), we extended the window forward and used leading data to calculate $Attn\_GeoPriceStDev_{ht}$.

Table 2 Results of the Seller Distribution Choice Model

	A: All sellers	B: Dealer sellers only	C: Dealer sellers only
Variable	Coef. (std. error)	Coef. (std. error)	Coef. (std. error)
β_1: $Propensity_{hikt}$	4.7179 (0.0644)***	4.3490 (0.0800)***	4.0677 (0.0802)***
β_2: $Propensity_{hikt} \times Attn\_Prevalence_{hkt}$	−1.6101 (0.1204)***	−1.4989 (0.1489)***	−1.2448 (0.1545)***
β_3: $Propensity_{hikt} \times Attn\_GeoPriceStDev_{ht}$	−0.0393 (0.0275)	−0.0503 (0.0343)	−0.0584 (0.0342)
β_4: $Supply_{hit}$	0.0856 (0.0044)***	0.1210 (0.0087)***	0.1253 (0.0090)***
β_5: $Supply^2_{hit}$	−0.0017 (0.0001)***	−0.0023 (0.0003)***	−0.0026 (0.0003)***
β_6: $PctSold_{hit}$	0.2591 (0.0209)***	0.2937 (0.0299)***	0.2858 (0.0304)***
β_7: $Price_{hit}$	0.0665 (0.0476)	0.0891 (0.0659)	0.0677 (0.0651)
β_{8i}: $Location(i)$	Included	Included	Included
β_9: $Distance_{gi}$	—	—	−0.1518 (0.0040)***
n (number of choices)	50,000	24,811	24,811
Log-likelihood	−53,688	−20,084	−19,126

Notes. This table presents the results of the seller distribution choice model. The dependent variable is $Choice_{ghikt}$, with $Choice_{ghikt} = 1$ if seller k chooses location i as the selling location for vehicle g and 0 otherwise. The variables shown in the table are explanatory variables that influence the seller's choice. See the appendix for variable definitions. To estimate the model, we randomly sampled 50,000 choices and then estimated the model using a conditional logit specification. We repeated this 100 times. Results shown are from the first subsample. The first column shows the results for all sellers, the second column shows the results when limited to dealer sellers, and the third column shows the results when limited to dealer sellers with $Distance_{gi}$ in the specification.

***Indicates significance at the 0.001 level.

represents the utility (disutility) associated with distributing vehicles to nearby (distant) locations. Vehicles' locations prior to entering the market are not recorded in the data; as a result, we do not include $Distance_{gi}$ in the utility function. However, we were able to estimate these locations and thereby calculate $Distance_{gi}$ for a subsample of the data, which allowed us to examine whether the omission of $Distance_{gi}$ affects our estimates. We did this as follows. Recall that there are two main types of sellers in the market: institutional sellers and dealer sellers (see §3.1). If a dealer who sells in the market also buys in the market, then we observe the zip code of his dealership, because the buyer's zip code is recorded in the data with each transaction. We assume that dealers store vehicles at their dealerships prior to offering them in the market. This assumption allowed us to calculate $Distance_{gi}$ for vehicles offered by dealer sellers as the number of miles between the dealer's zip code and the zip code of each location i, scaled by dividing by 100. Using this procedure, we were able to estimate the original location of approximately 50% of the vehicles in the data. We had no way of inferring the original location of vehicles offered by institutional sellers.

4.2. Estimation and Results

The data contain 53,691,888 choices (one per vehicle auctioned), and the mean number of alternatives per choice set is 12.1. This yields an overall choice data set of more than 650 million rows. The sheer size of the choice data set coupled with the nonlinearity of choice models makes model estimation difficult. To overcome this, we randomly sampled choices from the data and estimated the model using these subsamples. Each subsample consisted of 50,000 choices, and we estimated the model on 100 different subsamples using a conditional logit specification. Column A of Table 2 shows the results from the first subsample. Columns B and C show the results from the first subsample after limiting the choices to vehicles offered by dealer sellers; column C shows the results after including $Distance_{gi}$ in the specification. To assess whether the results reported in Table 2 are representative of the full data, we computed the mean and standard deviation of each coefficient across the 100 subsamples. We used the mean as the coefficient estimate and the standard deviation as a bootstrapped standard error. These results are consistent with those shown in Table 2 and are available upon request.

We first note that the coefficient estimates are similar regardless of whether we fit the model for all sellers (column A in Table 2) or for only dealer sellers (column B). Further, results are mostly unaffected by including $Distance_{gi}$ in the model for dealer sellers (column C), although the magnitudes of β_1 and β_2 are slightly reduced. This makes sense given that $Propensity_{hikt}$ is likely to be influenced by geographic proximity to a seller's vehicles, such that including $Distance_{gi}$ in the model should alter its effect. The consistency of our estimates indicates that exclusion of $Distance_{gi}$ from the model is unlikely to affect our coefficient estimates for the other variables.[9]

[9] Although the exclusion of $Distance_{gi}$ does not appreciably affect the coefficients of interest, it does affect the alternative-specific constants. This is because $Distance_{gi}$ accounts for some of the utility of each location otherwise reflected in the alternative-specific constants.

Table 3 Log-Likelihood Values for Different Versions of the Seller Distribution Choice Model

		After withholding				
	Full model	$Propensity_{hikt}$ and interactions	$Supply_{hit}$ and $Supply^2_{hit}$	$PctSold_{hit}$	$Price_{hit}$	$Distance_{gi}$
Log-likelihood (all sellers)	−53,688	−75,507	−53,894	−53,764	−53,689	—
Log-likelihood (dealer sellers only)	−19,126	−29,102	−19,238	−19,170	−19,127	−20,084

Notes. This table presents the log-likelihood values for the full seller distribution choice model and for versions in which variables are withheld. The first row corresponds to the model using all sellers. The second row corresponds to the model using only dealer sellers with $Distance_{gi}$ in the specification.

The β_1 coefficient is positive and significant, indicating that sellers offer vehicles of year/model h based on where they have offered vehicles of year/model h in the past. This provides evidence that habit plays a role in vehicle distribution. The β_2 coefficient is negative and significant, indicating that prior distribution choices play less of a role for vehicles with a high value of $Attn\_Prevalence_{hkt}$. This indicates that sellers are more likely to break with habit for vehicles that are prevalent in their fleet. We interpret this as evidence of more careful distribution of these vehicles. Consistent with this interpretation, we show below that $Attn\_Prevalence_{hkt}$ is negatively associated with the probability that a vehicle will be arbitraged (see §5.2). The β_3 coefficient is negative but insignificant. This suggests that, on average, a seller pays more attention to how prevalent a vehicle of year/model h is in her fleet than to how much its price has varied across locations when making distribution decisions. Given the insignificance of β_3 and β_7, we have no evidence that sellers consider historical prices when making distribution decisions (at least not on average).

In nonlinear models, the coefficient estimates on interaction terms do not necessarily indicate the direction of their marginal effects (Ai and Norton 2003).[10] To check the validity of our interpretation of the $Propensity_{hikt} \times Attn\_Prevalence_{hkt}$ coefficient, we estimated marginal effects via simulation as described below.

4.2.1. Economic Interpretation. We compared the log-likelihood values across different versions of the model to assess which variables have the most explanatory power. Table 3 shows the log-likelihood values for the full model and for versions of the model in which variables have been withheld. These values are based on estimation of the first subsample, but the results are consistent across subsamples.

Log-likelihood ("LL") ratio tests (i.e., $-2 \times (\mathrm{LL}_{\text{full}} - \mathrm{LL}_{\text{constrained}}) \sim \chi^2_{(\#\text{ of withheld variables})}$) confirm that withholding any of the variables except $Price_{hit}$ significantly harms model fit. However, the loss of fit is much more pronounced after withholding the $Propensity_{hikt}$ variables (including the interaction terms) than for any of the other variables, indicating that $Propensity_{hikt}$ is the major factor explaining sellers' choices in the model. We estimated the economic significance (i.e., the marginal effects) of $Propensity_{hikt}$ and $Propensity_{hikt} \times Attn\_Prevalence_{hkt}$ by using the results of the choice model shown in column A of Table 2 to simulate the increase in the number of vehicles sellers would send to a location if their $Propensity_{hikt}$ for that location were doubled. We also simulated how much this increase would be attenuated if $Attn\_Prevalence_{hkt}$ for the vehicle year/model being offered were doubled. We conducted these simulations for each of the 93 locations in the data. Doubling $Propensity_{hikt}$ for a location increases the number of vehicles offered at that location by an average of 20.1% (std. dev. 10.3%), but doubling the $Attn\_Prevalence_{hkt}$ for the vehicle year/model being offered attenuates this increase by an average of 8.7% (std. dev. 11.3%). We replicated the simulations using the results shown in column C of Table 2 to see how the estimates were affected by including $Distance_{gi}$ in the model. In this case, doubling $Propensity_{hikt}$ increases the number of vehicles offered at a location by an average of 11.1% (std. dev. 5.0%), but doubling the $Attn\_Prevalence_{hkt}$ for the vehicle year/model being offered attenuates this increase by an average of 7.5% (std. dev. 6.7%).

5. What Factors Affect Spatial Arbitrage Activity?

We first consider the factors, including seller attention allocation and electronic trading, that influence

[10] The interaction effect of two variables x_1 and x_2 is the cross-partial derivative of the expected value of the dependent variable with respect to x_1 and x_2 (Ai and Norton 2003). In a conditional logit model with $V_i = \beta_1 x_{1i} + \beta_2 x_{1i} x_{2i}$ (such as ours), the partial derivative of the probability P_i of choice i with respect to x_{1i} is $\partial P_i/\partial x_{1i} = (\beta_1 + \beta_2 x_{2i}) P_i (1 - P_i)$ (Train 2009), and the cross-partial derivative with respect to x_{1i} and x_{2i} is $\partial^2 P_i/(\partial x_{1i} \partial x_{2i}) = \beta_2 P_i (1 - P_i) + (\beta_1 + \beta_2 x_{2i}) \beta_2 x_{2i} (2P_i^2 - P_i)(P_i - 1)$. Using our estimates of β_1 and β_2 shown in column A of Table 2, $\partial^2 P_i/(\partial x_{1i} \partial x_{2i})$ is almost always negative for our model, with the exceptions occurring at unusually large values of x_{2i} and P_i.

Table 4 Descriptive Statistics for Logistic Regression Model for Whether a Purchased Vehicle Is Spatially Arbitraged

			Correlations											
	Mean	Std. dev.	1	2	3	4	5	6	7	8	9	10	11	12
1. $Arbitraged_g$	0.01	0.08	1											
2. $Attn\_Prevalence_{hkt}$	0.12	0.19	−0.01	1										
3. $Attn\_GeoPriceStDev_{ht}$[a]	1.85	0.92	0.00	−0.03	1									
4. $TotalVehicles\_Seller_{kt}$[b]	4.75	9.78	−0.01	−0.16	0.03	1								
5. $Webcast\_SecondHighestBid_g$	0.05	0.21	−0.01	−0.04	0.07	0.07	1							
6. $Webcast\_Buyer_g$	0.07	0.26	−0.02	−0.02	0.11	0.11	−0.07	1						
7. $NormalizedPrice_g$	0.99	0.21	−0.03	−0.01	0.01	−0.01	0.00	−0.02	1					
8. $Valuation_g$[c]	1.09	0.75	0.00	−0.02	0.67	0.13	0.10	0.16	−0.03	1				
9. $Mileage_g$[c]	5.70	4.29	0.01	0.04	−0.31	−0.26	−0.10	−0.15	0.05	−0.56	1			
10. $PctArbitraged\_Buyer_{mt}$[d]	0.70	3.50	0.42	−0.02	0.01	−0.01	−0.00	−0.04	−0.09	−0.00	0.01	1		
11. $PctArbitraged\_Seller_{kt}$[d]	0.69	0.88	0.11	−0.07	−0.03	0.01	−0.00	−0.05	−0.12	−0.05	0.06	0.13	1	
12. Day_t[a]	1.16	0.70	−0.01	−0.02	0.06	−0.02	0.12	0.14	−0.02	0.02	0.08	−0.00	−0.00	1

Note. See the appendix for variable definitions.
[a]Scaled by dividing by 1,000.
[b]Scaled by dividing by 100.
[c]Scaled by dividing by 10,000.
[d]Measured as percentages; e.g., 0.70 is 0.70%.

whether a vehicle is spatially arbitraged. We then consider the factors that influence arbitrageurs' choice of location at which to complete the arbitrage flip.

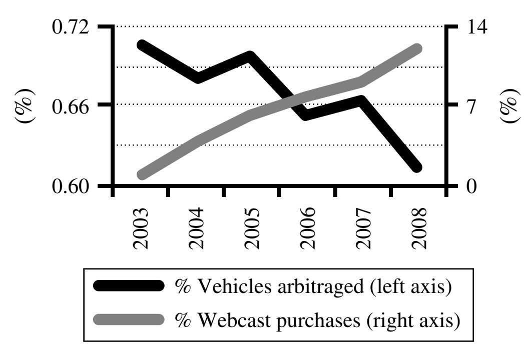
5.1. Specification of Logistic Regression Model for the Probability That a Vehicle Is Spatially Arbitraged

Each vehicle g that is sold will either be arbitraged ($Arbitraged_g = 1$) or not ($Arbitraged_g = 0$). We used logistic regression to examine which factors influence the probability that a sold vehicle is arbitraged, including how sellers allocate their attention across the vehicles in their inventory and the effect of electronic trading. We discuss these and the other variables in the logistic regression model below. Descriptive statistics for variables in the logistic regression model are provided in Table 4.

5.1.1. Seller Attention Allocation. The results of the seller distribution choice model show that (a) sellers are likely to offer vehicles of year/model h at the location where they have traditionally offered them, but (b) this probability is attenuated by $Attn\_Prevalence_{hkt}$. Habitual distribution of vehicles to the same location is likely to be suboptimal because it ignores market developments that cause the optimal selling location for vehicles to shift over time. This may result in a vehicle being offered at a location at which it fetches a price below that available at other locations, thereby creating an opportunity for an arbitrageur to flip it at a different location. The influence of habit decreases as $Attn\_Prevalence_{hkt}$ increases, suggesting that sellers pay more careful attention to the selling location of vehicles with high $Attn\_Prevalence_{hkt}$. We expect that this increased attention will decrease the probability that a vehicle of year/model h is arbitraged. We test this by including $Attn\_Prevalence_{hkt}$ in the logistic regression specification. We also include $Attn\_GeoPriceStDev_{ht}$ in the specification.

5.1.2. The Effect of Electronic Trading. As discussed in §3.1, the primary means by which buyers purchase vehicles electronically is the webcast channel. Figure 2 shows that the percentage of webcast purchases increased approximately linearly from 0.8% in 2003 to 11.9% in 2008. (We do not report 2009 data in Figure 2 because we do not observe the full year, and partial year statistics may be biased by seasonal effects.) Over the same time period, the percentage of vehicles arbitraged decreased from 0.71% to 0.61%. The correlation between the two trends is −0.89, which suggests that the webcast channel reduces opportunities for spatial arbitrage. We posit that the reason for the reduction in spatial arbitrage

Figure 2 Spatial Arbitrage and Webcast Purchase Activity by Year



Notes. This figure presents the percentage of vehicles spatially arbitraged and the percentage of vehicles purchased via the webcast channel by year. Arbitraged vehicles are those that were flipped across locations within seven days (i.e., $\alpha = 7$).

is that a buyer at a location where prices are high can use the webcast channel to buy a vehicle directly from a location where prices are low, rather than relying on an arbitrageur to buy from a low-priced location and to move the vehicle to him. Essentially, the webcast channel allows the buyer to disintermediate the arbitrageur. Because of this disintermediation potential, we expect that webcast bidding activity on a vehicle will decrease the probability that it is later arbitraged. We test this by including $Webcast\_Buyer_g$ and $Webcast\_SecondHighestBid_g$ in the logistic regression model. As discussed in §3.2, $Webcast\_Buyer_g$ and $Webcast\_SecondHighestBid_g$ indicate whether the high bid and second-highest bid for sold vehicles, respectively, were placed by a buyer using the webcast channel.[11]

5.1.3. Seller Characteristics. We included $TotalVehicles\_Seller_{kt}$ (scaled by dividing by 100) in the specification to control for effects attributable to the overall volume of vehicles offered by seller k at week t. We also included $PctArbitraged\_Seller_{kt}$, which is the number of vehicles sold by seller k that were arbitraged (using $\alpha = 7$) divided by the total number of vehicles sold by seller k, both over the 52 weeks (or fewer for observations in year 2003) prior to week t, scaled by multipliying by 100. This variable controls for unmodeled seller characteristics that might make them prone to having their vehicles arbitraged. We controlled for seller type by including four dummy variables derived from the *SellerType* variable. We used "factory" as the base case.

5.1.4. Vehicle Characteristics. We included $NormalizedPrice_g$, which is $Price_g$ divided by $Valuation_g$, because the more a vehicle costs relative to its valuation, the less likely a arbitrageur can turn a profit on it. We also controlled for $Valuation_g$, although we had no a priori expectations of its coefficient. 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.

5.1.5. Buyer Characteristics. $PctArbitraged\_Buyer_{mt}$ is the number of arbitrage flips (using $\alpha = 7$) completed by buyer m divided by the number of purchases made by buyer m, both over the 52 weeks (or fewer for observations in year 2003) prior to week t, scaled by multiplying by 100. This controls for the propensity of the buyer to engage in spatial arbitrage.

5.1.6. Location Characteristics. $Location(i)_g$, with $i = [2, \ldots, 93]$, are 92 dummy variables representing the location at which the transaction occurred. This controls for the possibility that arbitrage is more likely to originate at certain locations, 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.

5.1.7. Time. Day_t is the day the transaction occurs, which ranges from 1 (January 1, 2003) to 2,318 (May 7, 2009), scaled by dividing by 1,000. It controls for unmodeled variables that change over time that affect the likelihood of arbitrage, such as learning by traders in the market.

5.2. Estimation and Results

We set $Arbitraged_g = 1$ if a vehicle g was flipped at a different location within α days of the original transaction (i.e., $DaysToFlip \leq \alpha$). In our main results, we set $\alpha = 7$, but used different values of α for robustness. Because estimating a series of logistic regressions for different values of α 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. We used clustered standard errors (clustered by location and day) to account for unobserved location-specific and time-specific factors that may create correlation among error terms. 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 are shown in Table 5.

The coefficient for $Attn\_Prevalence_{hkt}$ ($\beta = -0.347$) is negative and significant. A one standard deviation increase in $Attn\_Prevalence_{hkt}$ is associated with a 6.4% decrease in the odds that a vehicle is arbitraged. $Attn\_Prevalence_{hkt}$ decreases the odds that a vehicle is

[11] We are assuming that webcast bidders would not have otherwise been physical bidders if the webcast channel was not available, in other words, that bidders use the webcast channel to participate in auctions in which they otherwise would not have participated. We examined this assumption by using panel regression with fixed effects 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 he purchased. The specification is $\#Locations_{mt} = \beta_0 + \beta_1 \#WebcastPurchases_{mt} + \beta_2 \#TotalPurchases_{mt} + c_m + \sum_{t=2003}^{2008} \beta_t Year(t) + \varepsilon_{mt}$, where m indexes the buyer and t the year, $\#TotalPurchases_{mt}$ controls for each buyer's overall purchase volume, c_m represents a buyer fixed effect, and $Year(t)$ are dummy variables for each year, which are included to control for a possible time trend. The R^2 statistic is 0.75. The β_1 coefficient is positive and significant ($\beta_1 = 0.02$, robust standard error $= 0.002$), and a one standard deviation increase in $\#WebCastPurchases_{mt}$ is associated with a 27% increase in $\#Locations_{mt}$. This indicates that buyers use the webcast channel to purchase from locations from which they would not have purchased physically.

Table 5 Results of Logistic Regression Model for Whether a Purchased Vehicle Is Spatially Arbitraged

Variable	Coef. (std. error)	Relationship to odds of vehicle being arbitraged
$Attn\_Prevalence_{hkt}$	−0.347 (0.019)***	A one standard deviation increase associated with a 6.4% decrease in the odds of the vehicle being arbitraged.
$Attn\_GeoPrice\ StDev_{ht}$	−0.002 (0.001)***	A one standard deviation increase associated with a 0.2% decrease in the odds of the vehicle being arbitraged.
$TotalVehicles\_Seller_{kt}$	0.006 (0.000)***	A one standard deviation increase associated with a 5.7% increase in the odds of the vehicle being arbitraged.
$SecondHighestBid\_Webcast_g$	−0.098 (0.016)***	$Webcast\_SecondHighestBid_g = 1$ associated with a 9.3% decrease in the odds of the vehicle being arbitraged.
$Webcast\_Buyer_g$	−0.758 (0.027)***	$\text{Webcast\_Buyer}_g = 1$ associated with a 53.1% decrease in the odds of the vehicle being arbitraged.
$NormalizedPrice_g$	−0.921 (0.016)***	A one standard deviation increase associated with a 17.6% decrease in the odds of the vehicle being arbitraged.
$Valuation_g$	0.474 (0.023)***	Increase from \$10,000 to \$20,000 associated with a 26.9% increase in the odds of the vehicle being arbitraged. Inflection point in curvilinear relationship reached at approximately \$30,200.
$Valuation_g^2$	−0.078 (0.006)***	
$Mileage_g$	0.107 (0.004)***	Increase from 10,000 to 20,000 associated with a 10.0% increase in the odds of the vehicle being arbitraged. Inflection point in curvilinear relationship reached at approximately 143,800.
$Mileage_g^2$	−0.004 (0.000)***	
$PctArbitraged\_Buyer_{mt}$	0.159 (0.000)***	A one standard deviation increase associated with a 74.5% increase in the odds of the vehicle being arbitraged.
$PctArbitraged\_Seller_{kt}$	0.238 (0.003)***	A one standard deviation increase associated with a 23.2% increase in the odds of the vehicle being arbitraged.
Day_t	−0.155 (0.007)***	Increase from day 1 to day 2,318 associated with a 30.2% decrease in the odds of the vehicle being arbitraged.
$SellerType(2)_g$	0.520 (0.020)***	*SellerType* = "Lease" associated with a 68.2% increase in the odds of the vehicle being arbitraged compared to *SellerType* = "Factory."
$SellerType(3)_g$	0.407 (0.032)***	*SellerType* = "Rental" associated with a 50.2% increase in the odds of the vehicle being arbitraged compared to *SellerType* = "Factory."
$SellerType(4)_g$	0.563 (0.023)***	*SellerType* = "Repossession" associated with a 75.6% increase in the odds of the vehicle being arbitraged compared to *SellerType* = "Factory."
$SellerType(5)_g$	0.401 (0.021)***	*SellerType* = "Dealer" associated with a 49.3% increase in the odds of the vehicle being arbitraged compared to *SellerType* = "Factory."
Constant	−5.886 (0.049)***	n/a
92 location dummies		Included

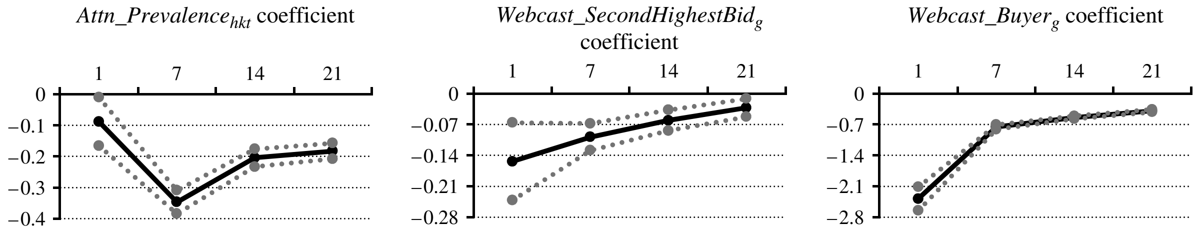
Notes. This table presents the results of the logistic regression model of the probability that a purchased vehicle is spatially arbitraged. The dependent variable is $Arbitraged_g$, with $Arbitraged_g = 1$ if the vehicle was arbitraged and 0 otherwise. The variables shown in the table are explanatory variables that influence the probability of arbitrage. See the appendix for variable definitions. Pseudo-R^2: 0.35; log-pseudolikelihood = −763,318; $n = 31,805,961$.

***Indicates significance at the 0.001 level.

arbitraged by as much as 29% (i.e., $1-\exp(-0.347)$). This provides evidence that the more attention a seller pays to a vehicle's distribution, the lower the probability that the vehicle will be arbitraged. The coefficient for $Attn\_GeoPriceStDev_{ht}$ is also negative and significant, although this may be an artifact of sample size, as the economic significance of the coefficient is negligible. The coefficient for $Webcast\_SecondHighestBid_g$ is negative and significant; $Webcast\_SecondHighestBid_g = 1$ decreases the odds that a vehicle is arbitraged by 9.3%. This indicates that webcast bidding activity lowers the probability of spatial arbitrage. The coefficient for $Webcast\_Buyer_g$ is also negative and significant; $Webcast\_Buyer_g = 1$ decreases the odds that a vehicle is arbitraged by 53.1%. Although some of this effect may reflect an underlying propensity for buyers who purchase vehicles via the webcast channel to *not* engage in spatial arbitrage, some of the effect likely represents how the webcast channel eliminates spatial arbitrage activity by allowing buyers to disintermediate arbitrageurs, particularly because $PctArbitraged\_Buyer_{mt}$ is controlled for.

5.2.1. Robustness Checks. To assess whether our results were sensitive to the threshold we used for delineating arbitrage flips from other types of flips, we reran the logistic regression model for different values of α. Figure 3 plots the coefficients and 95% confidence intervals for $Attn\_Prevalence_{hkt}$, $Webcast\_SecondHighestBid_g$, and $Webcast\_Buyer_g$ at $\alpha = 1$, $\alpha = 7$, $\alpha = 14$, and $\alpha = 21$. The $Attn\_Prevalence_{hkt}$ coefficient is negative and significant at all levels of α. The coefficients for the variables capturing webcast bidding activity, $Webcast\_SecondHighestBid_g$ and $Webcast\_Buyer_g$, are negative and significant at all levels of α, although they become less negative at increasing levels of α. We draw three conclusions from these robustness checks. First, the results are robust to different definitions of what constitutes spatial arbitrage. Second, the webcast bidding activity effects attenuate at higher levels of α, indicating that they only affect flips that are completed quickly. This

Figure 3 Robustness of the Logistic Regression Results to Alternative Values of the α Threshold Used to Delineate Spatial Arbitrage



Notes. The key results of the logistic regression model for whether a vehicle is spatially arbitraged (see Table 5) might be sensitive to the α threshold we used to delineate spatial arbitrage. To test this, we reran the model using different values of the α threshold (1, 7, 14, and 21). This figure shows coefficients (black dots) and 95% confidence intervals (gray dots) for the $Attn\_Prevalence_{hkt}$, $Webcast\_SecondHighestBid_g$, and $Webcast\_Buyer_g$ variables from the logistic regression model estimated using different values of the α threshold (1, 7, 14, and 21). The coefficients and confidence interval values are shown on the y-axis and the values of the α threshold are shown on the x-axis.

makes sense given that bidding activity 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 arbitrageurs 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.

5.3. Specification of Arbitrageur Distribution Choice Model

For every vehicle that is spatially arbitraged ($n = 210{,}170$), the arbitrageur l must choose a destination location j. We used a discrete choice model to examine the variables affecting the arbitrageur's choice of a destination location. We defined each arbitrageur's choice set analogously to how we defined each seller's choice set in §4.1, except we removed alternatives in which the potential destination location j was the same as the source location i. We modeled the utility of each alternative location j as $U_{ghijlt} = V_{ghijlt} + \varepsilon_{ghijlt}$, where g is the vehicle, h is the vehicle's year/model, i is the source location, j is the destination location, l is the arbitrageur, t is the week, and $V_{ghijkt} = \beta_1 \times Propensity_{hjlt} + \beta_2 \times Propensity_{hjlt} \times Attn\_Prevalence_{hlt} + \beta_3 \times Propensity_{hjlt} \times Attn\_GeoPriceStDev_{ht} + \beta_4 \times Distance_{ij} + \beta_5 \times Distance_{ij} \times Webcast\_SecondHighestBid_g + \beta_6 \times Supply_{hjt} + \beta_7 \times Supply^2_{hjt} + \beta_8 \times PctSold_{hjt} + \beta_9 \times Price_{hjt} + \sum_{j=2}^{93} \beta_{10(j)} \times Location(j)$.

Our model for how arbitrageurs choose selling locations for vehicles mirrors that for how sellers choose selling locations (note the change in subscripts), except that we include $Distance_{ij}$ and $Distance_{ij} \times Webcast\_SecondHighestBid_g$ in the specification. $Distance_{ij}$ is the distance in miles between the source location i and the potential destination location j, scaled by dividing by 100. We interact $Distance_{ij}$ with $Webcast\_SecondHighestBid_g$ to investigate how arbitrageurs' sensitivity to distance when flipping vehicles is affected by webcast bidding activity at the source location. Our motivation for this is as follows. We first assume that arbitrageurs (a) prefer to purchase vehicles in low-priced locations and resell 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 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 arbitrageurs, forcing them to transport vehicles to more distant destination location to complete the flips.[12] To illustrate, assume that pickup trucks trade at low prices in Dallas but at high prices in Houston. This will create an incentive for 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 arbitrageurs. Table 6 provides descriptive statistics for the arbitrageur choice model.

[12] We did not interact $Distance_{ij}$ with $WebCast\_Buyer_g$ because $WebCast\_Buyer_g$ does not have the same interpretation in this model as in the logistic regression model. In the logistic regression model, which includes all transactions, $WebCast\_Buyer_g$ represents whether the buyer (be they arbitrageur or otherwise) purchased the vehicle via the webcast channel. In the arbitrageur choice model, which only includes arbitrage flips, $WebCast\_Buyer_g$ represents whether the *arbitrageur* purchased the vehicle via the webcast channel. Because we have no a priori expectation on how the arbitrageur's buying channel affects his choice of destination location, we did not include $WebCast\_Buyer_g$ in the model. (Incidentally, if we include $WebCast\_Buyer_g$ by interacting it with $Distance_{ij}$, its coefficient is 0.0803 (std. error = 0.0077) with the other coefficients essentially unchanged.)

Table 6 Descriptive Statistics for the Arbitrageur Distribution Choice Model

	Mean	Std. dev.	Correlations 1	2	3	4	5	6	7	8
1. $Propensity_{hjlt}$	0.07	0.24	1							
2. $Attn\_Prevalence_{hlt}$	0.06	0.12	0.02	1						
3. $Attn\_GeoPriceStDev_{hlt}$[a]	1.93	0.78	−0.01	−0.02	1					
4. $Distance_{ij}$[b]	5.30	4.83	−0.11	−0.12	0.03	1				
5. $Webcast\_SecondHighestBid_g$	0.04	0.20	0.00	−0.00	0.04	−0.00	1			
6. $Supply_{hjt}$[c]	0.89	2.40	0.16	0.04	0.00	−0.04	−0.00	1		
7. $PctSold_{hjt}$	0.43	0.39	0.10	0.04	−0.06	−0.02	−0.01	0.19	1	
8. $Price_{hjt}$[d]	1.16	0.64	0.01	−0.02	0.61	0.05	0.06	0.06	−0.10	1

Note. See the appendix for variable definitions.
[a]Scaled by dividing by 1,000.
[b]Scaled by dividing by 100.
[c]Scaled by dividing by 10.
[d]Scaled by dividing by 10,000.

5.4. Estimation and Results

We estimated the arbitrageur distribution choice model using a conditional logit specification. Results appear in Table 7. The $Distance_{ij}$ coefficient is negative and significant, indicating that sellers prefer nearby locations over distant locations, ceteris paribus. However, the $Distance_{ij} \times Webcast\_SecondHighestBid_g$ coefficient is positive and significant, indicating that webcast bidding activity causes the arbitrageur to be less sensitive to distance and thereby to choose a more distant destination location. This indicates that webcast activity is not only making spatial arbitrage less common (see §5.2), but it is also making it more difficult. Similar to §4.2, we checked the validity of our interpretation of the $Distance_{ij} \times Webcast\_SecondHighestBid_g$ coefficient and examined the economic significance of $Distance_{ij}$ and $Distance_{ij} \times Webcast\_SecondHighestBid_g$ by simulating the decrease in the number of times arbitrageurs choose a destination location if the location were 50% farther away from the source location. We also simulated how much this decrease would be attenuated if $Webcast\_SecondHighestBid_g = 1$. We conducted these simulations for each location in the data. Results show that arbitrageurs would choose a location j on 15.5% (std. dev. 12.8%) fewer occasions if $Distance_{ij}$ is increased by 50%, but that this decrease is attenuated by 12.7% (std. dev. 11.3%) if $Webcast\_SecondHighestBid_g = 1$.

Table 7 Results of Arbitrageur Distribution Choice Model

Variable	Coef. (std. error)
β_1: $Propensity_{hjlt}$	4.8499 (0.0414)***
β_2: $Propensity_{hjlt} \times Attn\_Prevalence_{hlt}$	−3.1792 (0.0995)***
β_3: $Propensity_{hjlt} \times GeoPriceStDev_{hlt}$	−0.0436 (0.0194)*
β_4: $Distance_{ij}$	−0.2106 (0.0017)***
β_5: $Distance_{ij} \times Webcast\_SecondHighestBid_g$	0.0253 (0.0066)***
β_6: $Supply_{hjt}$	0.0754 (0.0033)***
β_7: $Supply^2_{hjt}$	−0.0017 (0.0001)***
β_8: $PctSold_{hjt}$	0.4100 (0.0108)***
β_9: $Price_{hjt}$	0.4864 (0.0270)***
$\beta_{10,i}$: $Location(i)$	Included

Notes. This table presents the results of the arbitrageur distribution choice model. The dependent variable is $Choice_{ghijlt}$, with $Choice_{ghijlt} = 1$ if arbitrageur l chooses location j as the destination location at which to complete the arbitrage flip and 0 otherwise. The variables shown in the table are explanatory variables that influence the arbitrageur's choice. See the appendix for variable definitions. Log-likelihood = −151,897; n (number of choices) = 210,710.

*Indicates significance at the 0.05 level; ***indicates significance at the 0.001 level.

Comparing the results shown in Tables 2 and 7 provide interesting insights into how sellers and arbitrageurs distribute vehicles. First, the direction and significance of β_1 and β_2 are the same in both tables, indicating that *Propensity* and *Attn_Prevalence* affect the choices of sellers and arbitrageurs in the same directions. However, we replicated the simulations we conducted in §4.2.1 and identified some behavioral distinctions. First, doubling *Propensity* for arbitrageurs increases the number of vehicles offered at a location by an average of 9.9% (std. dev. 8.1%), compared to an average of 20.1% (std. dev. 10.3%) for sellers (or 11.1% (std. dev. 5.0%) if we limit the sample to dealer sellers and include $Distance_{gi}$ in the seller distribution choice model). Second, doubling $Attn\_Prevalence_{hkt}$ for the vehicle year/model being offered attenuates this increase by an average of 14.3% (std. dev. 13.9%) for arbitrageurs, compared to an average of 8.7% (std. dev. 11.3%) for sellers (or 7.5% (std. dev. 7.7%) if we limit the sample to dealer sellers and include $Distance_{gi}$ in the seller distribution choice model). This indicates that habit plays less of a role in arbitrageurs' decisions than in sellers' decisions. Second, *Price* has a significant and positive effect on arbitrageurs' choices, but no effect on sellers' choices. Put together, these two findings indicate that arbitrageurs pay greater

Table 8 Count and Percentage of Arbitrage Flips Based on Our Estimates of Whether the Arbitrageur Profited and the Seller Forwent Additional Revenue

	Seller forgoes additional revenue	Seller does not forgo additional revenue	Total
Arbitrageur profits	181,306 (86.3%)	0 (0.0%)	181,306 (86.3%)
Arbitrageur does not profit	19,069 (9.1%)	9,795 (4.7%)	28,864 (13.7%)
Total	200,375 (95.3%)	9,795 (4.7%)	210,170

Notes. The northwest cell of the matrix shows the number of flips for which the arbitrageur profits while the seller forgoes additional revenue. The northeast cell shows the number of flips for which the arbitrageur profits while the seller does not forgo revenue. The southwest cell shows the number of flips for which the arbitrageur *does not* profit while the seller forgoes revenue. This can occur if the arbitrageur's fees are higher than those of the seller. The southeast cell shows the number of flips for which the arbitrageur does not profit while the seller does not forgo revenue. Arbitraged vehicles are those that were flipped across locations within seven days (i.e., $\alpha = 7$).

attention to recent pricing trends and are more flexible in where they distribute vehicles than sellers. This makes sense given the nature of the arbitrageurs' business model.

6. How Large and Variable Are Spatial Arbitrage Profits, and How Much Revenue Are Sellers Forgoing?

We leveraged the granularity and detail of the data to develop rough estimates of the profitability of spatial arbitrage. We also estimated the additional revenue that sellers are forgoing by "misdistributing" vehicles that are later arbitraged. For each arbitrage flip, we calculated the following values. We calculated the arbitrageur's gross revenue as the price of the vehicle at the destination location. We calculated the arbitrageur's cost as the sum of the price of the vehicle at the source location, the buy fee at the source location, the estimated cost to transport the vehicle from the source location to the destination location, the estimated cost of the capital tied up in the vehicle while being arbitraged, and the sell fee at the destination location. We calculated the arbitrageur's profit by subtracting the arbitrageur's estimated cost from his gross revenue. We calculated the seller's net revenue as the price of the vehicle at the source location minus the sell fee at the source location. We estimated what we refer to as the seller's forgone net revenue as the price of the vehicle at the destination location minus the estimated cost to transport the vehicle from the source location to the destination location, the estimated cost of deferring liquidating the vehicle while it is being transported to and sold at the destination location, and the estimated sell fee at the destination location.[13] We calculated the difference between our estimate of the seller's forgone net revenue and her actual net revenue. If this difference is positive, then we say that the seller "forwent additional revenue" by not selling the vehicle directly at the destination location.

Table 8 shows a 2×2 matrix in which the rows represent whether the arbitrageur profited or not and the columns represent whether the seller forwent additional revenue or not, based on our estimates. Table 8 shows the results for $\alpha = 7$; increasing or decreasing the α threshold has minimal impact on the distribution of flips across cells in the matrix.

We calculated total arbitrage gross revenue by summing the price at the destination location for each flip, and we estimated total arbitrage profits/losses by summing the profit/loss estimate for each flip. We also summed our estimates of the additional revenue forgone by sellers; we did not include flips in which the seller did not forgo revenue in this calculation. These estimates are shown in Table 9 for the entire time span covered by the data, on a per year basis, and on a per vehicle basis for different thresholds of α.

The total amounts of arbitrage gross revenue, arbitrage profit, and additional revenue forgone by sellers naturally increase with the number of flips classified as arbitrage (as determined by α), but the estimates per vehicle are relatively stable. Arbitrageurs incur costs of approximately $10,700 to receive an average

[13] We estimated the transportation cost by using a transportation quote system provided by a major auction company to generate quotes between 50 zip code pairs chosen randomly from the zip codes of the locations in our data. The relationship between the miles between zip codes and the quotes is fit well ($R^2 = 0.99$) by a second-degree polynomial regression model, which we used to estimate the transport cost between all pairs of locations in our data. We adjusted our transportation cost estimates by the average U.S. diesel price per gallon (U.S. Energy Information Administration 2011) for the month in which the arbitrage flip was initiated to allow for changes in fuel prices over time. We measured the cost of capital for arbitrageurs using the prime rate $+\gamma$ to represent the interest incurred while holding the vehicle. We measured the cost of capital for sellers using the risk-free rate $+\lambda$ to represent the interest deferred by not liquidating the vehicle. We varied both γ and λ for sensitivity purposes, which does not affect our results because of the limited number of days for interest to accrue. We assumed that the sell fee paid by the seller is the same at the source and destination locations.

Table 9 Estimates of Arbitrage Gross Revenues, Arbitrage Profits, and Additional Revenue Forgone by Sellers for Different Levels of the Arbitrage Threshold α

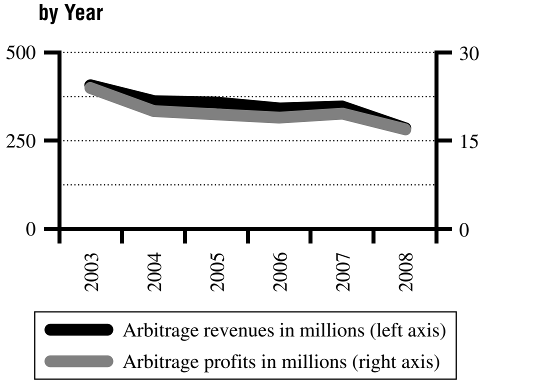
		Arbitrage gross revenues			Arbitrage profits			Additional revenue forgone by sellers		
α	Number of arbitrage flips	Total[a]	Per year[a]	Per vehicle[b]	Total[a]	Per year[a]	Per vehicle[b]	Total[a]	Per year[a]	Per vehicle[b]
1	21,709	230.0	36.2	10,595	12.8	2.0	593	21.5	3.4	1,043
3	59,301	660.6	104.1	11,140	37.4	5.9	630	61.0	9.6	1,077
7	**210,170**	**2,376.3**	**374.3**	**11,307**	**135.3**	**21.3**	**644**	**221.9**	**35.0**	**1,108**
14	393,613	4,453.2	701.5	11,314	248.5	39.2	631	415.2	65.4	1,127
21	515,819	5,804.1	914.3	11,252	315.8	49.7	612	540.3	85.1	1,139
28	599,321	6,706.3	1,056.5	11,190	356.1	56.1	594	623.2	98.2	1,149

Notes. This table shows estimates of arbitrage gross revenues, arbitrage profits, and the additional revenue forgone by sellers for different levels of the α threshold used to delineate spatial arbitrage. Estimates are provided for the entire time span (shown as the "total" column), on a per year basis, and on a per vehicle basis. We highlight the estimates at $\alpha = 7$ as our focal estimates.
[a]In millions of dollars.
[b]In raw dollars.

profit of $600 per arbitraged vehicle (standard deviation approximately $800). This represents an average accounting profit of approximately 5.6%. As discussed in §3.3, 50% of the spatial arbitrage transactions are conducted by 0.06% of the buyers in the data, and 80% of arbitrage transactions are conducted by 0.38% of the buyers. The arbitrageurs who account for 50% of the arbitrage transactions average 130 flips per year, which represents an average annual profit of approximately $78,000 per arbitrageur. The ten most active arbitrageurs account for 15% of the arbitrage transactions and average 496 flips per year each; these arbitrageurs achieve estimated annual profits of $311,000 each. We also investigated arbitrage gross revenues and estimated profits over time. Figure 4 shows that both revenues and profits are decreasing over time. Both results are consistent with our findings that spatial arbitrage has become less common and more difficult as electronic trading has become more common.

For instances in which sellers forgo additional revenue, the average amount is approximately $1,100. For any given seller, this is mostly inconsequential. To illustrate, consider a seller who sells 150 vehicles. At an average price of $10,000 per vehicle, this yields total revenue of $1.5 million. At $\alpha = 7$, approximately one of these vehicles (0.7%) will be arbitraged, costing the seller $1,100, or 0.07% of her total revenue. If we set α more liberally at $\alpha = 28$, the additional forgone revenue still represents only 0.22% of the seller's total revenue. However, the total amount of additional revenue forgone by the sellers is substantial and has created a healthy industry for the arbitrageurs, who receive positive returns on approximately $374 million in annual transaction revenues (at $\alpha = 7$).

Figure 4 Estimated Spatial Arbitrage Gross Revenues and Profits by Year



Notes. Spatial arbitrage gross revenues and estimate profits per year. Arbitraged vehicles are those that were flipped across locations within seven days (i.e., $\alpha = 7$).

Our analysis allows us to examine two of the claims about arbitrage offered by Shleifer and Vishny (1997). First, the evidence that the vast majority of spatial arbitrage is done by a small fraction of traders provides empirical support for the claim that relatively few traders engage in arbitrage. Second, the $10,700 average incurred cost, the variance in profits, and the result that arbitrageurs take a loss approximately 14% of the time (see Table 8) provide empirical support for Shleifer and Vishny's (1997) argument that arbitrage requires capital and is risky.

7. Conclusion

7.1. Intended Contributions

Despite the fundamental role that arbitrage plays in finance and economic theory, there is relatively little empirical evidence of the factors that create and eliminate arbitrage opportunities, how often arbitrage occurs, and how profitable arbitrage transactions are. The goal of this study was to address these gaps by empirically investigating spatial arbitrage in the wholesale automotive industry. We contribute to the literature on arbitrage in three main ways.

First, we examine the behavioral factors that lead to arbitrage opportunities. A spatial arbitrage opportunity arises when a seller sells a product at a location where prices are below those available at another location, after accounting for transportation and related transaction costs. We used a discrete choice model to examine how sellers choose selling locations. We find that habit plays an important role, such that sellers tend to sell vehicles where they have sold similar vehicles in the past. However, sellers are more likely to break with habit for vehicles that are prevalent in their fleet. We interpret this as evidence that sellers pay more attention to how they distribute these vehicles, and we find that these vehicles are less likely to be arbitraged. Although it is possible that sellers are behaving irrationally by offering a portion of their vehicles at suboptimal locations, the revenue that sellers forgo by doing so is modest relative to their overall revenues, suggesting that it may be rational for sellers to accept these losses rather than to commit the resources to erase them.

Second, we conclude that information technology has improved market efficiency by reducing the frictions associated with geographic trade. Our findings indicate that electronic trading decreases spatial arbitrage opportunities by giving buyers the reach to purchase directly from remote locations, rather than relying on arbitrageurs to purchase from these locations and to move the vehicles to them. Thus, electronic trading appears to undermine a traditional function of the arbitrageurs, which has been to serve as "market makers" who provide liquidity by taking positions on vehicles and matching buyers and sellers across locations. This is consistent with findings reported in Hendershott and Moulton (2011), who found that floor broker and specialist trading volumes on the New York Stock Exchange declined significantly after the introduction of the OpenBook electronic order system.

Information technology improves seller outcomes in our context, although it is worth noting the mechanism by which this occurs. Prior research has suggested that information technology might help market participants expand their information processing capabilities (e.g., Einav et al. 2009, Peng and Xiong 2006). In our context, this would improve sellers' ability to optimize vehicle distribution. Although information technology may be helping sellers in this way, the mechanism we document is different. We show that even if sellers continue to make suboptimal distribution decisions, information technology can make those decisions less costly by bringing remote bidders into the auctions at the suboptimal locations.

Third, the granularity and detail of our data allows us to develop estimates of the prevalence and value of arbitrage and to test propositions about how arbitrage operates in practice. Most prior empirical research on arbitrage has inferred arbitrage activity based on price data. Our data, which contain unique identifiers for individual assets, traders, and locations, allow us to observe individual arbitrage transactions with a higher level of precision. We find support for two of the claims about arbitrage proposed by Shleifer and Vishny (1997). First, most traders do not engage in arbitrage: over 90% of arbitrage transactions are conducted by less than 1% of the traders. Second, arbitrage requires capital and is risky. Arbitrageurs incur costs of approximately \$10,700 per arbitraged vehicle. Although they average \$600 profit per arbitraged vehicle, profits are variable (standard deviation $\sim$ \$800) and arbitrageurs take a loss approximately 14% of the time. Overall, we estimate that approximately 1% of sold vehicles are arbitraged, and that sellers often forgo over \$1,000 in additional revenue for each vehicle arbitraged. This is a modest amount for any individual seller given her total revenues, but creates a robust industry for arbitrageurs when aggregated across sellers. However, the arbitrage industry is shrinking as electronic trading improves market efficiency and erodes the opportunities for arbitrageurs.

7.2. Limitations and Future Research

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 arbitrage. For example, 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 is an opportunity for future research.

Although our analysis is specific to the wholesale automotive market, we believe 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 seller distribution decisions are likely to create pricing disparities across locations for spatial 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 spatial arbitrage at a transaction level. Similar identifiers are available in other markets and will become increasingly available as advanced identification and tracking technologies become more widely adopted. Future research might use methods similar to ours to estimate the incidence and economic value of spatial arbitrage in other markets and the factors that affect arbitrage opportunities.

Acknowledgments

The authors thank Brad Barber (special issue coeditor), the associate editor, and three anonymous reviewers for helping them to improve the paper. The authors also thank seminar participants at the Georgia Institute of Technology, New York University, the 2010 Statistical Challenges in Electronic Commerce Research Symposium, and the 2010 International Conference on Information Systems.

Appendix. Definition of Variables

Variable	Description
Variables in the raw data	
Sold?	Indicator variable for whether the vehicle was sold.
SellerID	Unique identifier for the seller in the transaction.
VIN	Vehicle Identification Number: Unique identifier for the vehicle in the transaction.
Date	Date the vehicle was auctioned.
VehicleModel	Make and model of the vehicle.
VehicleYear	Model year of the vehicle.
Mileage	Odometer reading of the vehicle.
LocationID	Unique identifier for the location at which the vehicle was auctioned.
LocationZip	Five-digit zip code for the location at which the vehicle was auctioned.
SellerType	Set of indicator variables representing the type of seller auctioning the vehicle.
BuyerID	Unique identifier for the buyer in the transaction.
BuyerZip	Five-digit zip code for the buyer in the transaction.
BuyFee	Fee paid by the buyer in the transaction.
SellFee	Fee paid by the seller in the transaction.
Price	Transaction price.
Valuation	Estimated wholesale price of the vehicle in the transaction. Calculated by the auction company based on transactions for similar vehicles over the prior 30 days.
Webcast_Buyer	Indicator variable for whether the vehicle was purchased by a buyer using the webcast channel.
SecondHighBid_Webcast	Indicator variable for whether the second-highest bid was placed by a buyer using the webcast channel.
Location(*i*)	Set of indicator variables for each location i at which vehicles are auctioned.
Variables constructed from the raw data	
DaysToFlip	Number of days between the two transactions comprising an arbitrage flip.
$Propensity_{hikt}$	Number of vehicles of year/model h that seller k offered at location i divided by the number of vehicles of year/model h that seller k offered at all locations, both in the 52 weeks (or fewer for observations in year 2003) prior to week t.
$Attn\_Prevalence_{hkt}$	Percentage of vehicles of year/model h offered by seller k in week t relative to all vehicles offered by seller k in week t.
$Attn\_GeoPriceStDev_{ht}$	Standard deviation of the mean prices of vehicles of year/model h across locations. Calculated using prices from the three weeks prior to week t.
$Supply_{hit}$	Number of vehicles of year/model h offered at a location i in the three weeks prior to week t.
$PctSold_{hit}$	Percentage of vehicles of year/model h offered at location i in the three weeks prior to week t that were sold.
$Price_{hit}$	Average price of vehicles of year/model h sold at location i over the three weeks prior to week t.
$Distance_{gi}$	Distance in miles between a vehicle g's location and location i.
$TotalVehicles\_Seller_{kt}$	Number of vehicles offered by seller k in week t.
$NormalizedPrice_g$	The transaction price of a vehicle divided by its valuation estimate.
$PctArbitraged\_Seller_{kt}$	Number of vehicles sold by seller k that were arbitraged divided by the total number of vehicles sold by seller k, both over the 52 weeks (or fewer for observations in year 2003) prior to week t.
$PctArbitraged\_Buyer_{mt}$	Number of arbitrage flips completed by buyer m divided by the number of purchases made by buyer m, both over the 52 weeks (or fewer for observations in year 2003) prior to week t.
Day_t	Day on which vehicle was auctioned. Ranges from 1 (1/1/2003) to 2,318 (5/7/2009).
$Arbitraged_g$	Indicator variable set to 1 if a purchased vehicle was later arbitraged and 0 otherwise.

References

Aarts, H., B. Verplanken, A. van Knippenberg. 1998. Predicting behavior from actions in the past: Repeated decision making or a matter of habit? *J. Appl. Soc. Psych.* **28**(15) 1355–1374.

Ai, C., E. Norton. 2003. Interaction terms in logit and probit models. *Econom. Lett.* **80**(1) 123–129.

Alexander, C., J. Wyeth. 1994. Cointegration and market integration: An application to the Indonesian rice market. *J. Development Stud.* **30**(2) 303–334.

Barber, B., T. Odean. 2008. The effect of attention on the buying behavior of individual and institutional investors. *Rev. Financial Stud.* **21**(2) 785–818.

Barrett, C. 2008. Spatial market integration. S. Durlauf, L. Blume, eds. *The New Palgrave Dictionary of Economics*, 2nd ed. Palgrave Macmillan, London.

Baulch, B. 1997. Transfer costs, spatial arbitrage, and testing for food market integration. *Amer. J. Agricultural Econom.* **79**(2) 477–487.

Brown, J., A. Goolsbee. 2002. Does the Internet make markets more competitive? Evidence from the life insurance industry. *J. Political Econom.* **110**(3) 481–507.

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

Corwin, S., J. Coughenour. 2008. Limited attention and the allocation of effort in securities trading. *J. Finance* **63**(6) 3031–3067.

Cox, D. 1972. Regression models and life tables. *J. Royal Statist. Soc. B. Methodological* **34**(2) 187–202.

Einav, L., M. Jenkins, J. Levin. 2009. The impact of information technology on consumer lending. Working paper, Stanford University, Stanford, CA. http://www.stanford.edu/~leinav/Credit_Scoring.pdf.

Fackler, P., B. Goodwin. 2001. Spatial price analysis. B. Gardner, G. Rausser, eds. *Handbook of Agricultural Economics*, Vol. 1B. Elsevier, Amsterdam, 971–1024.

Hendershott, T., P. C. Moulton. 2011. Automation, speed, and stock market quality: The NYSE's Hybrid. *J. Financial Markets* **14**(4) 568–604.

Jensen, R. 2007. The digital provide: Information (technology), market performance, and welfare in the South Indian fisheries sector. *Quart. J. Econom.* **122**(3) 879–924.

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

Klemperer, P. 1987. Markets with consumer switching costs. *Quart. J. Econom.* **102**(2) 375–394.

Mitchell, M., T. Pulvino. 2001. Characteristics of risk and return in risk arbitrage. *J. Finance* **56**(6) 2135–2175.

Mitchell, M., L. Pedersen, T. Pulvino. 2007. Slow moving capital. *Amer. Econom. Rev.* **97**(2) 215–220.

National Auto Auction Association. 2011. About Us. Accessed December 23, http://www.naaa.com/about_us/about_us.html.

National Automobile Dealers Association (NADA). 2009. NADA data 2009. Report, NADA, McLean, VA. http://www.nada.org/nadadata.

Overby, E., C. Forman. 2011. The effect of electronic commerce on geographic trade and price variance in a business-to-business market. Working Paper 11-30. NET Institute, New York. http://ssrn.com/abstract=1958056.

Peng, L., W. Xiong. 2006. Investor attention, overconfidence and category learning. *J. Finance Econom.* **80**(3) 563–602.

Pontiff, J. 1996. Costly arbitrage: evidence from closed-end funds. *Quart. J. Econom.* **111**(4) 1135–1151.

Ravallion, M. 1986. Testing market integration. *Amer. J. Agricultural Econom.* **68**(1) 102–109.

Samuelson, W., R. Zeckhauser. 1988. Status quo bias in decision making. *J. Risk Uncertainty* **1**(1) 7–59.

Schiraldi, P. 2011. Automobile replacement: A dynamic structural approach. *RAND J. Econom.* **42**(2) 266–291.

Sharpe, W., G. Alexander, J. Bailey. 1995. *Investments*. 5th ed. Prentice Hall, Upper Saddle River, NJ.

Shleifer, A., R. Vishny. 1997. The limits of arbitrage. *J. Finance* **52**(1) 35–55.

Simon, H. 1955. A behavioral model of rational choice. *Quart. J. Econom.* **69**(1) 99–118.

Train, K. 2009. *Discrete Choice Methods with Simulation*, 2nd ed. Cambridge University Press, New York.

U.S. Energy Information Administration. 2011. Short-term energy outlook. Accessed March 1, http://explore.data.gov/Energy-and-Utilities/Short-Term-Energy-Outlook-Real-Petroleum-Prices/zinz-q5tn.