informs
http://pubsonline.informs.org/journal/isre/

INFORMATION SYSTEMS RESEARCH
Vol. 28, No. 1, March 2017, pp. 97–116
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Electronic Commerce, Spatial Arbitrage, and Market Efficiency

Hemang Subramanian,[a] Eric Overby[b]

[a] Florida International University, Miami, Florida 33196; [b] Georgia Institute of Technology, Atlanta, Georgia 30308
Contact: hsubrama@fiu.edu (HS); eric.overby@scheller.gatech.edu (EO)

Received: April 4, 2014
Revised: May 9, 2015; February 11, 2016
Accepted: May 24, 2016
Published Online in Articles in Advance: November 3, 2016

https://doi.org/10.1287/isre.2016.0653

Abstract. Electronic commerce can improve market efficiency by helping buyers and sellers find and transact with each other across geographic distance. We study the effect of two distinct forms of electronic commerce on market efficiency, which we measure via the existence and exploitation of spatial arbitrage opportunities. Spatial arbitrage represents a more precise measure of market efficiency than does price dispersion, which is typically used, because it accounts for the transaction costs of trading across distance and for unobserved product heterogeneity. One of the forms of electronic commerce that we study is a webcast channel that allows electronic access to the traditional physical market, while the other is a standalone electronic market. Both forms provide traders with expanded reach to find and transact with each other across geographic distance, but only one allows traders to conduct transactions immediately, at any time. This variance helps us isolate the effect of these mechanisms (reach and transaction immediacy) on efficiency. We find that electronic commerce reduces the number of arbitrage opportunities (likely by expanding traders' geographic reach) but improves arbitrageurs' ability to identify and exploit those that remain (likely by expanding arbitrageurs' reach and providing them with the immediacy to exploit opportunities quickly). Overall, our results suggest that electronic commerce improves market efficiency not only by helping buyers and sellers transact across distance (thereby balancing supply and demand across geographic locations) but also by helping arbitrageurs quickly exploit any remaining arbitrage opportunities (thereby rebalancing supply and demand across geographic locations).

History: Sanjeev Dewan, Senior Editor; Michael Zhang, Associate Editor.
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2016.0653.

Keywords: electronic commerce • spatial arbitrage • market efficiency • quasi-natural experiment • coarsened exact matching • automotive

1. Introduction

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

Addressing this question allows us to make the following contributions. First, scholars typically use price dispersion to measure market efficiency. By contrast, we use the existence and exploitation of spatial arbitrage opportunities, which is more precise for markets in which trading is geographically distributed. Following Coleman (2009), we define spatial arbitrage as the purchase and subsequent resale of the same product in different geographic locations to exploit a price discrepancy.[1] If buyers and sellers do not match efficiently across locations, then this creates spatial arbitrage opportunities in which a third party—an arbitrageur—purchases products from sellers in locations where prices are low and resells them to buyers in locations where prices are high. Spatial arbitrage will occur as long as the transaction costs associated with moving products between locations are lower than the price difference between locations, and it will become more prevalent as the gap between the transaction costs and the price difference widens. This measure has several advantages over price dispersion, including inherently accounting for transaction costs and for unobserved product heterogeneity. To illustrate the first advantage, consider a perfectly efficient market, i.e., one in which buyers and sellers match optimally, regardless of geographic location. According to the law of one price, this market might have substantial price dispersion, because prices in an efficient market can vary up to the transaction costs of moving products between locations (Persson 2008). Yet there would be no spatial arbitrage in this market; indeed, "no arbitrage" is a classic condition for market efficiency (Barrett 2008, Takayama and Judge 1971). Thus, a researcher using price dispersion to measure market efficiency might incorrectly conclude that this market was inefficient. A researcher using spatial arbitrage to measure market efficiency would not make this mistake. Use of this new measure is important because improving measurement is fundamental to scientific advancement.

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

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

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

Theoretically, the expanded reach provided by both channels should help "regular" buyers purchase directly from sellers in remote locations, thereby disintermediating the arbitrageurs and reducing arbitrage opportunities. Yet it should also help arbitrageurs find and exploit previously hidden opportunities. The transaction immediacy provided by the standalone electronic market should also help arbitrageurs identify and exploit arbitrage opportunities before they dissipate. This suggests a nuanced effect in which electronic commerce eliminates many arbitrage opportunities but improves arbitrageurs' ability to identify and exploit those that remain. Whether each of the channels increases or decreases the number of arbitrage transactions (i.e., exploited opportunities) depends on whether the channel increases the efficiency with which arbitrageurs exploit opportunities more than it reduces the number of opportunities.

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

Overall, our results provide a fuller and more nuanced picture of the way in which electronic commerce affects market efficiency than has previously been documented (to our knowledge). Per the law of one price (e.g., Persson 2008), imbalances in supply and demand are likely to occur in all geographically distributed markets at times, creating spatial arbitrage opportunities. If a market is efficient, then arbitrage opportunities will be few (because supply and demand are mostly balanced across locations) and will not persist for long (because any remaining supply/demand imbalances will be quickly exploited by arbitrageurs, thereby restoring balance). Our analysis shows that electronic commerce improves efficiency through both mechanisms. First, electronic commerce helps buyers and sellers trade with each other across geographic locations, thereby providing better supply/demand balance and eliminating arbitrage opportunities. This is consistent with prior research (e.g., Aker 2010, Jensen 2007). Second, electronic commerce helps arbitrageurs exploit the remaining supply/demand imbalances, thereby returning the market to efficiency when it strays. This mechanism is more nuanced and has not been empirically shown to our knowledge.

2. Literature Review and Differentiators of Our Study

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

2.1. Research on Market Efficiency and Electronic Commerce

Scholars have typically measured market efficiency in geographically distributed markets by examining price dispersion (e.g., Aker 2010, Jensen 2007). The intuition is that a high degree of price dispersion indicates that products are not being allocated efficiently, i.e., that supply is too high for the demand in some regions and too low in other regions. This will lead to low prices in the former regions and to high prices in the latter regions, creating price dispersion. Scholars have also measured market efficiency by analyzing the comovement of prices at different locations over time, which is essentially an analysis of how price dispersion evolves longitudinally (Alexander and Wyeth 1994). The intuition for this measure is that prices in an efficient market can differ across locations up to the cost of transport between locations (Persson 2008, Takayama and Judge 1971), but this difference should be relatively constant over time, i.e., prices across locations should move up or down together, such that price dispersion remains mostly unchanged. If prices do not move together, then this might reflect excess supply or demand in some locations at certain times, reflecting the market's inefficiency at balancing supply and demand.

To assess the effect of electronic commerce on market efficiency, scholars have examined whether price dispersion is lower online than off-line (Brown and Goolsbee 2002, Brynjolfsson and Smith 2000). A common motivation for these studies is that because transaction costs are lower online than off-line (as is typically assumed), supply and demand will be more efficiently distributed online, resulting in lower price dispersion online. Empirical support for this is mixed (Baye et al. 2006). Another approach scholars have used is to examine whether price dispersion in a market declines as electronic commerce is adopted. Studies using this approach have shown that electronic commerce has led to reduced price dispersion across locations (e.g., Aker 2010, Overby and Forman 2015, Parker et al. 2016).

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

2.1.1. Measuring Efficiency via the Existence and Exploitation of Spatial Arbitrage Opportunities Rather Than via Price Dispersion.

Despite its widespread use, price dispersion has two key limitations as a measure of efficiency in geographically distributed markets (Badiane and Shively 1998, Barrett 2008). The first limitation is that a "baseline" level of price dispersion may exist in a perfectly efficient market, with this baseline equal to the transaction costs of moving products between locations (Takayama and Judge 1971). Without knowledge of these costs, it is difficult to determine what level of price dispersion represents the baseline level versus what level represents inefficiency in matching buyers and sellers. Prior research that has relied on price dispersion to measure efficiency has often worked around this issue by studying price dispersion on a relative basis, e.g., if the level of price dispersion in a market declines over time, then it is reasonable to assume that efficiency has improved. However, this approach makes the (potentially erroneous) assumption that the baseline does not also decline.[2] To illustrate this limitation, consider (hypothetically) that the price of a barrel of crude oil is $45 in the United States and $55 in the United Kingdom. On one hand, this $10 price dispersion might reflect an inefficient imbalance in supply and demand, such that buyers in the United Kingdom (sellers in the United States) would be better off shifting their demand (supply) to the United States (United Kingdom) to exploit the inefficiency. On the other hand, supply and demand might be perfectly balanced, with the $10 dispersion representing the baseline level and resulting from different tax/tariff regimes and/or nonzero transport costs between the United States and the United Kingdom. In this case, there is no inefficiency for buyers or sellers to exploit.

The second limitation of price dispersion as a measure of efficiency is that it will be inaccurate if products whose prices are being compared are not comparable because of unobserved factors such as differences in quality. Although this is not a concern for perfectly homogeneous products, it is a concern for any product whose attributes may vary from unit to unit (e.g., agricultural crops, automobiles, crude oil, metals, etc.). To illustrate this limitation, consider that crude oil in the United Kingdom might be of higher quality than that in the United States. In this case, the $10 price dispersion would reflect this quality difference, not inefficiency in balancing supply and demand.

Using the existence and exploitation of spatial arbitrage opportunities as the measure of market efficiency addresses these two limitations. First, the arbitrage measure inherently accounts for the transaction costs of moving products between locations. To elaborate, consider that although it is difficult for researchers to assess whether a given level of price dispersion exceeds the transaction costs of moving products between locations, it is far less difficult for arbitrageurs. This is because arbitrageurs are market specialists who are keenly aware of price disparities and the transaction costs associated with exploiting them (Shleifer and Vishny 1997). Arbitrageurs consider the transaction costs when determining whether to engage in arbitrage. If the transaction costs are lower than the price disparities, then arbitrage will occur, becoming more prevalent as the gap between the transaction costs and the price disparities widens, i.e., as the market becomes more inefficient. If the transaction costs are higher than the price disparities, then arbitrage will not occur. Essentially, using spatial arbitrage to measure efficiency removes the burden from the researcher of trying to estimate whether a given level of price dispersion represents inefficiency. Instead, this question is answered (directly) by the arbitrageurs. Second, the arbitrage measure is immune to potential quality differences across products, because the same product is traded in both locations (unless the product is altered during transport).

To illustrate these advantages, we return to the crude oil example. If we observe arbitrage activity in which arbitrageurs purchase in the United States for $45 per barrel and resell in the United Kingdom for $55 per barrel, then we can safely assume that the transaction costs are less than $10 per barrel and that the market is inefficient. Also, because the same oil is being transacted in both locations, we can eliminate the possibility that the $10 price dispersion is due to quality differences. On the other hand, if we do not observe arbitrage activity, then the $10 price dispersion is more likely to reflect nonzero transaction costs or differences in quality than inefficiency. Table 1 summarizes the advantages of measuring market efficiency via the existence and exploitation of spatial arbitrage opportunities instead of via price dispersion.

2.1.2. Examining Two Distinct Forms of Electronic Commerce. A limitation of research on electronic commerce and market efficiency is that it tends to implicitly assume that all forms of electronic commerce are the same. Prior studies have either examined a single form of electronic commerce and used that to draw conclusions about electronic commerce in general (e.g., Jensen 2007) or studied electronic commerce in general terms without differentiating between different types of electronic commerce (e.g., Brynjolfsson and Smith 2000). A differentiator of our study is that we examine the effects of two distinct forms of electronic commerce: a webcast channel that permits electronic access to the traditional physical market and a standalone electronic market. This helps us examine the mechanisms responsible for the effect of electronic commerce. To elaborate, different forms of electronic commerce have distinct features. Assume that form 1 of electronic commerce has feature A and no effect on market efficiency, while form 2 has features A and B and a positive effect on market efficiency. This would suggest that feature B (and not feature A) is the reason for the effect. Such a

Table 1. Issues with Measuring Market Efficiency via Price Dispersion and How Those Are Remedied by Using the Existence and Exploitation of Spatial Arbitrage Opportunities as the Measure

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

Table 2. Categorization of Research on the Effect of Electronic Commerce on Market Efficiency, Including Differentiators of the Present Study

		Measure of market efficiency	
		Price dispersion	Spatial arbitrage
Analysis of different forms of electronic commerce?	Yes		*Present study*
	No	Aker (2010), Brown and Goolsbee (2002), Chellappa et al. (2011), Jensen (2007), etc.	Overby and Clarke (2012)

conclusion would not be possible if only one of these two forms of electronic commerce were examined. Our analysis of multiple forms of electronic commerce also contributes to the growing literature that recognizes that different types of electronic commerce (e.g., desktop computer versus phone based) may have distinct effects (e.g., Ghose et al. 2013). Table 2 summarizes how our research is distinct from prior research on the effect of electronic commerce on market efficiency.

2.2. Research on Spatial Arbitrage and Arbitrageur Behavior

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

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

3. Theory and Hypotheses

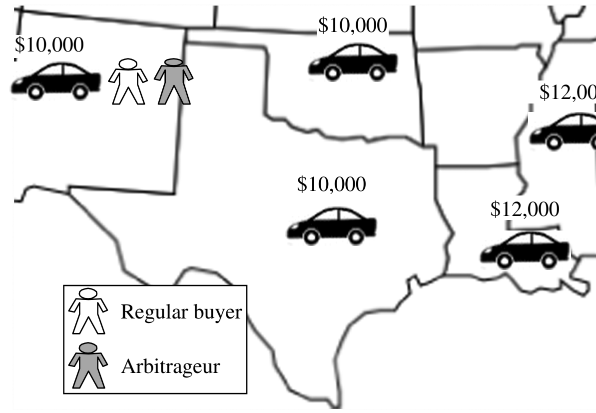
3.1. Effect of Electronic Commerce on Spatial Arbitrage Opportunities

To consider the theoretical effect of electronic commerce on spatial arbitrage opportunities, it is useful to consider the features that distinguish electronic commerce from traditional commerce. We focus on two features for our analysis: (a) reach, which is a spatial feature, and (b) transaction immediacy, which is a temporal feature. First, we define reach as the ability for traders to find and transact with each other across geographic distance. Electronic commerce expands reach by lowering traders' search costs, such that they can more easily find and consummate trading opportunities with partners in remote locations. Second, electronic commerce typically provides transaction immediacy, which we define as the ability to conduct a transaction immediately, at any time (e.g., when physical stores are closed, without having to wait in line, before the price changes, etc.).[3]

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

The transaction immediacy of electronic commerce should also affect spatial arbitrage. Arbitrageurs are continuously looking for opportunities in which products in location k are priced below their value in location l. These opportunities are fluid and do not persist indefinitely, because prices depend on dynamic supply/demand conditions in both locations. Thus, if an arbitrageur determines that the price for a product at location k is sufficiently below what he expects to receive at location l, then it is important for him to exploit this arbitrage opportunity before conditions change. The reach provided by electronic commerce will help arbitrageurs find these opportunities, and the transaction immediacy of electronic commerce will allow arbitrageurs to purchase the products before their prices change or they are purchased by a "regular" buyer. Thus, the transaction immediacy of electronic commerce should enhance arbitrageurs' ability to exploit arbitrage opportunities. Regular buyers could also leverage the transaction immediacy of electronic commerce to purchase undervalued products, which would take opportunities away from the arbitrageurs. However, we expect arbitrageurs to take fuller advantage of transaction immediacy to purchase undervalued products. This is because there are many instances in which a product will appear undervalued

Figure 1. Illustration of Why More Products Will Appear Undervalued to Arbitrageurs Than to Regular Buyers



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

to an arbitrageur but not to a regular buyer. To see this, consider that a regular buyer is only interested in whether a product in a remote location is undervalued relative to his location, i.e., he goes online to get a better deal than he can locally. By contrast, arbitrageurs are interested in whether a product in *any* location is undervalued relative to *any other* location. Figure 1 provides an illustration. Also, research in behavioral economics suggests that arbitrageurs will identify and seize arbitrage opportunities more quickly than will regular buyers, because the former have the expertise and time to monitor market conditions more closely (Peng and Xiong 2006, Shleifer and Vishny 1997).

The logic above suggests that increasing use of electronic commerce should lead to a decrease in the number of arbitrage opportunities but to an increase in the percentage of arbitrage opportunities that are exploited. Both mechanisms reflect increased market efficiency. The first reflects better balance of supply/demand across geographic locations, and the second reflects better rebalancing of supply/demand when inefficiencies arise (as they inevitably do).

Hypotheses 1A, B (H1A, B). *Increasing use of electronic commerce is* (a) *negatively associated with the number of spatial arbitrage opportunities, but* (b) *positively associated with the percentage of spatial arbitrage opportunities that are exploited.*

Whether electronic commerce leads to an increase or decrease in the number of arbitrage opportunities that are exploited (i.e., the number of arbitrage transactions) depends on whether the increased efficiency with which arbitrageurs exploit opportunities (i.e., the "opportunity exploitation" effect) outpaces the reduction in opportunities (i.e., the "opportunity reduction" effect). This should depend on the features provided by electronic commerce. In particular, assuming that two forms of electronic commerce both provide expanded reach but only one provides transaction immediacy, the form that provides transaction immediacy is more likely to foster arbitrage transactions (or at least inhibit them less).

Hypothesis 2 (H2). *Different forms of electronic commerce will have different effects on the number of spatial arbitrage transactions, with forms that support transaction immediacy having a more positive effect.*

3.2. Factors That Affect Arbitrageur Behavior of Where to Source Products for Later Arbitrage

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

Hypotheses 3A, B (H3A, B). *Arbitrageurs will prefer source locations at which* (a) *a high percentage of products sell for below their market value, and* (b) *prices for the products being sold vary widely across locations.*

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

Hypotheses 4A, B (H4A, B). *Arbitrageurs will prefer source locations that are* (a) *nearby, but* (b) *relatively difficult for other traders to access.*

4. Empirical Context

The empirical context for our study is the U.S. wholesale used vehicle market. Buyers in this market are used vehicle dealers. Most use the market as a source of inventory for their retail lots, while a small minority use the market to source vehicles for arbitrage within the wholesale market. The former (whom we will refer to as "buyers" or "regular buyers") try to make money from the difference between vehicles' retail and wholesale prices. The latter (whom we will refer to as "arbitrageurs") try to make money by exploiting inefficiencies within the wholesale market. Sellers are of two types: (a) used vehicle dealers, and (b) institutional sellers such as rental car firms. The former use the market either to sell vehicles that they do not sell retail or (much less commonly) to sell vehicles that they are arbitraging within the wholesale market. Institutional sellers use the market to sell vehicles retired from their fleets. Data were provided by an intermediary in the market that facilitates trades between buyers and sellers. The intermediary operates over 70 physical market facilities across the United States as well as the webcast channel and standalone electronic market described in Section 4.1. The data consist of 40,657,724 successful transactions facilitated by the intermediary from 2003 to 2010. Variables are described in Table 3.

Traditionally, the U.S. wholesale used vehicle market has functioned as a physical market in which buyers, sellers, and vehicles are collocated at market facilities. Each facility has a large parking lot for vehicle storage and a warehouse-type building equipped with multiple *lanes*, which are essentially one-way streets. Transactions occur as follows. Vehicles are driven down a lane—one at a time—where buyers interested in that vehicle will have gathered. An auctioneer solicits bids for each vehicle and awards the vehicle to the highest bidder, assuming he meets the seller's reserve price. This process takes 30–45 seconds, after which another vehicle is auctioned. It is common for vehicles to be auctioned in multiple lanes at the same facility concurrently. A "presale" list of the vehicles to be offered is provided in advance on the intermediary's website. Buyers can link to vehicle information, photographs, and a condition report (if one was written for that vehicle). Buyers can also use the intermediary's website to search for where and when vehicles of interest will be offered, including whether they are available via the standalone electronic market (see Section 4.1).

4.1. Two Distinct Forms of Electronic Commerce

In the early 2000s, the intermediary who operates these facilities began simulcasting via the Internet the physical auctions as they were occurring at the facilities. This allows buyers to experience the live audio and video of the auctions via an Internet browser, and it permits them to bid on vehicles in competition with the buyers who are physically collocated at the facility. As such, this "webcast" channel provides buyers with electronic access to the auctions occurring in the physical market. The webcast channel was implemented in phases. This is because implementing the channel required that

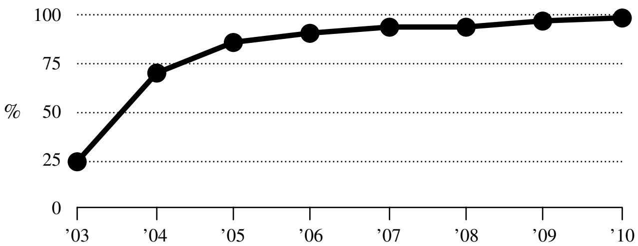
Table 3. Description of Variables

Variable	Description	Descriptive statistics
Arbitraged	Denotes whether a purchased vehicle was spatially arbitraged (= 1) or not (= 0).	Mean: 0.0079; SD: 0.0885
FacilityID	The ID for the facility at which the vehicle was located.	—
FacilityZip	The facility's zip code.	—
LaneID	The ID for the lane at the facility in which the vehicle was auctioned.	—
TraderID	The ID for the trader. The same ID is used regardless of whether the trader is the buyer or seller in a transaction.	—
BuyerZip	The buyer's zip code. Could also be thought of as *ArbitrageurZip* when the buyer later arbitrages the vehicle.	—
BuyerDistance	The distance in miles between *BuyerZip* and *FacilityZip*.	Mean: 205; SD: 343
RemoteBuyer	Denotes whether *BuyerDistance* was at least one standard deviation above the mean (= 1) or not (= 0).	Mean: 0.12; SD: 0.32
SellerType	Denotes the type of seller (institutional or dealer). 1 = Institutional; 0 = Dealer.	Mean: 0.57; SD: 0.50
SellerPctArbitraged	The percentage of vehicles sold by the seller over the previous 28 days that were arbitraged. If the seller sold 0 vehicles over the previous 28 days, we increased the number of days until the number sold was greater than 0.	Mean: 0.006; SD: 0.023
SaleDate	The date the vehicle was sold.	—
VIN	The vehicle's vehicle identification number.	—
VehicleYear	The model year of the vehicle.	Mean: 2006.6; SD: 2.43
Make	The make of the vehicle, e.g., Chevrolet, Toyota, etc.	—
Model	The model of the vehicle, e.g., Tahoe, Camry, etc.	—
Price	The vehicle's sales price.	Mean: 10,470; SD: 8,060
Valuation	The vehicle's market value as estimated by the intermediary that provided the data.	Mean: 10,589; SD: 8,051
Mileage	The odometer reading of the vehicle.	Mean: 64,302; SD: 49,473
StandaloneElectronicMarket	Denotes whether a vehicle was offered in the standalone electronic market (= 1) or the physical market (= 0).	Mean: 0.02; SD: 0.14
WebcastEnabled	Denotes whether a vehicle in the physical market was offered in a webcast enabled lane (= 1) or not (= 0).	Mean: 0.76; SD: 0.44
WebcastBuyer	Denotes whether a vehicle offered in the physical market was purchased by a buyer using the webcast channel (= 1) or the physical channel (= 0).	Mean: 0.09; SD: 0.28
BuyFee	The transaction fee paid by the buyer.	Mean: 182.1; SD: 144.9
SellFee	The transaction fee paid by the seller.	Mean: 134.5; SD: 184.6
TransportFee	The transport fee between the facility and the buyer's location. Not available for all transactions; see the online appendix.	Mean: 160.8; SD: 161.8
ListingDate	The date a vehicle was listed on the standalone electronic market.[a]	—
TransactionType	Denotes whether a vehicle was purchased in the standalone electronic market via the Buy Now mechanism or auction (either before or after the hidden reserve price is met).[a]	Buy Now: 58%; Auction (after reserve met): 35%; Auction (before reserve met): 7%

[a]Not available for all transactions; see Section 5.2.3.

camera, microphone, and other equipment be installed *in each lane at each facility*. This means that we observe many instances in which highly similar vehicles were auctioned at the same facility on the same day, some in lanes that were equipped for webcast and some in lanes that were not. As discussed in Section 5.2.1, we leverage this in a quasi-natural experiment to assess the effect of the webcast channel on the probability that a purchased vehicle is later arbitraged. Figure 2 shows how the percentage of vehicles available via webcast increased as the channel was deployed.[4] The intermediary also operates a standalone electronic market whose format is similar to that of eBay. In this market, sellers post listings of their vehicles, and buyers have the option to purchase them for a fixed "Buy Now" price or to bid for them over a period that spans several hours or a few days. Typically, a winning bidder exceeds the seller's (hidden) reserve price. Yet sometimes the seller will sell to the high bidder before the reserve price has been met.

Figure 2. Annual Percentage of Vehicles Offered in the Physical Market That Were Offered in Webcast Enabled Lanes



From a theoretical perspective, the webcast channel and the standalone electronic market differ in their ability to provide expanded reach and transaction immediacy. Both forms of electronic commerce provide expanded reach: using either form, buyers can easily search for and purchase vehicles from remote locations without having to travel.[5] However, because of their different designs, only the standalone electronic market provides transaction immediacy. This is because the webcast channel is an electronic access channel to auctions occurring in the physical market. As a result, the webcast channel can only be used to purchase vehicles during the 30–45 second window within which they are being auctioned at the physical market facility. Also, a buyer can only purchase a vehicle in the webcast channel if he places the highest bid and has this bid accepted by the seller. This means that the webcast channel cannot be used by a buyer to purchase a vehicle immediately, at any time; i.e., there is no transaction immediacy. By contrast, most vehicle listings in the standalone electronic market include a posted "Buy Now" price. Buyers can click this button at any time to purchase the vehicle immediately.

To explore the implications of these differences further, it is useful to consider what information arbitrageurs consider when determining whether they can profitably arbitrage a vehicle. This includes (a) a vehicle's price, (b) its location, and (c) its expected price at other locations a few days in the future. Neither the webcast channel nor the standalone electronic market provides any particular benefit to arbitrageurs for items (b) and (c) beyond other information sources such as the intermediary's website. This is because a vehicle's location is published on the intermediary's website, and arbitrageurs estimate expected prices at other locations based on historical transactions (available via the intermediary's website), their expectation of future supply/demand patterns (which may be informed by the presale lists available on the intermediary's website), etc. However, the two channels differ in their ability to provide arbitrageurs with information about a vehicle's price, particularly whether the vehicle is undervalued and thus a good candidate for arbitrage. Arbitrageurs using the standalone electronic market can see if a vehicle is undervalued based on its Buy Now price and (potentially) based on the current high bid for the vehicle and how long bidding will remain open. By contrast, arbitrageurs using the webcast channel do not know whether a vehicle will be undervalued until the 30–45 second auction for it occurs. Thus, the standalone electronic market is better suited for arbitrageurs to scan for undervalued vehicles, and it provides the transaction immediacy necessary to quickly exploit the associated arbitrage opportunities.

4.2. Identification of Spatial Arbitrage Transactions

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

5. Analysis and Results

5.1. Testing H1: The Association Between Electronic Commerce, the Number of Arbitrage Opportunities, and the Percentage of Opportunities That Are Exploited

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

Panel A of Figure 3 shows that the number of arbitrage opportunities decreased by 72% over our sample period, from an estimated 1,532,232 in 2003 to 426,122 in 2010. Panel B of Figure 3 shows that the number of arbitrage opportunities that were exploited (i.e., the number of arbitrage transactions) also decreased over time (by 46%). Panel D of Figure 3 shows that electronic trading via both channels increased over time, with most of the electronic trading occurring via the webcast channel. The correlations between the time series for the percentage of electronic trading and (a) the number of arbitrage opportunities, and (b) the number of arbitrage opportunities that were exploited (i.e., the number of arbitrage transactions) are −0.98 ($p < 0.01$) and −0.95 ($p < 0.01$).[8] This supports H1A. Panel C of Figure 3 shows that the *percentage* of arbitrage opportunities that were exploited increased by 93% over our sample period, from 2.4% to 4.7%. The correlation between this time series and that for the percentage of electronic trading is 0.96 ($p < 0.01$). This supports H1B and suggests that electronic commerce improved the ability for arbitrageurs to exploit inefficiencies, even as it decreased the overall number of inefficiencies.[9]

5.2. Testing H2: The Effect of Each Form of Electronic Commerce on the Number of Arbitrage Transactions

5.2.1. Testing the Effect of the Webcast Channel. We tested the effect of the webcast channel on spatial arbitrage by leveraging the phased adoption of the webcast channel to conduct a quasi-natural experiment (e.g., Jensen 2007, Overby and Forman 2015). Because the webcast technology was deployed at different times for each lane at each facility, we observe many instances in which similar vehicles were sold at the same facility on the same day, some in lanes that were webcast enabled and some in lanes that were not. We considered the former to be potential "treated" vehicles and the latter potential "control" vehicles. If whether a vehicle was sold in a webcast enabled lane was randomly assigned, then we could identify the treatment effect of the webcast channel via a simple comparison of outcomes for the treated and control vehicles. However, as with much observational data, this is not the case. When the webcast channel was deployed, sellers and facility managers (who collectively determine the lanes in which vehicles are offered) tended to use webcast enabled lanes for vehicles whose *VehicleYear*/*Make*/*Model* (e.g., 2002 Audi TT) was only available in a few facilities (so buyers from other areas

Figure 3. Spatial Arbitrage and Electronic Trading Trends Over Time

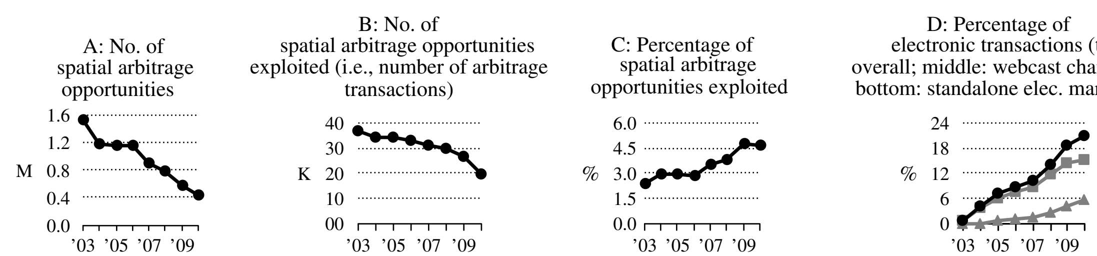


Table 4. Descriptive Statistics for Treated and Control Observations for Testing the Effect of the Webcast Channel, Before and After Matching

Variable	*n*: Strata	*n*: Treated vehicles	*n*: Control vehicles	Mean: Treated vehicles	Mean: Control vehicles	Difference in means	(*t*-stat)
			Panel A. Before matching (full sample)				
Valuation	n/a	15,892,484	9,058,903	11,841.36	8,521.38	3,319.98	(1,049.53)
Mileage	n/a	15,892,484	9,058,903	54,291.36	76,537.35	−22,245.99	(−1,047.43)
SellerPctArbitraged	n/a	15,892,484	9,058,903	0.0068	0.0067	0.0001	(5.12)
DayOfWeek	n/a	15,892,484	9,058,903	3.12	3.04	0.08	(201.49)
			Panel B. After matching (matched sample)				
Valuation	19,095	36,027	26,810	11,960.21	11,960.52	−0.31	(−0.25)
Mileage	19,095	36,027	26,810	38,633.18	38,636.40	−3.22	(1.26)
SellerPctArbitraged	19,095	36,027	26,810	0.00538	0.00530	0.00008	(2.01)
DayOfWeek	19,095	36,027	26,810	3.13	3.14	−0.01	(1.57)

Notes. Observations are from 2003 to 2007. The means of the variables in the matched sample differ from those in the full sample. However, we believe the matched sample is reasonably representative. This is because the 95% confidence intervals around *Mileage* and *Valuation* in the matched sample cover 94% of the transactions in our full sample, i.e., the matched sample contains many matches of not only low mileage, high value vehicles but also of high mileage, low value vehicles.

could bid for them electronically)[10] and that were relatively new with low mileage (so personal inspection of the vehicle was not required to assess quality). These assignment criteria became less relevant over time as the webcast channel was increasingly deployed. Nevertheless, vehicles sold in webcast and nonwebcast enabled lanes differed in *VehicleYear*, *Make*, *Model*, *Mileage*, and *Valuation* (see Table 4).

Although assignment to webcast enabled lanes was not random, because we know the factors that influenced assignment (*VehicleYear*, *Make, Model*, and *Mileage*), we control for them. To do this (and to control for other factors), we matched control vehicles to treated vehicles on *VehicleYear*, *Make*, *Model*, *Mileage*, *Valuation*, *FacilityID*, *SellerType, SellerPctArbitraged*, and *SaleDate*. This increases the likelihood that the only material difference between the control and treated vehicles is that the latter were sold on webcast enabled lanes. Essentially, the matched control vehicles serve as counterfactuals for what would have happened to the treated vehicles if they had not been treated, thereby allowing us to estimate the treatment effect of being sold in a webcast enabled lane on whether a vehicle is arbitraged (see Iacus et al. 2012). Matching on *FacilityID* is a key part of our identification strategy. In many cases, this allows us to compare vehicles sold in a webcast enabled lane to very similar vehicles sold in a nonenabled lane just a few feet away. Matching on *SellerPctArbitraged* is also important, because this controls for a host of unobserved seller characteristics that influence the probability that vehicles they sell are arbitraged. One such characteristic is how effective sellers are at distributing vehicles to the "right" locations so that they do not sell for below-market discounts (becoming good candidates for arbitrage).

We matched vehicles using exact matching and coarsened exact matching ("CEM"). Each treated vehicle was only matched to a control vehicle with the same *FacilityID*, *VehicleYear*, *Make*, *Model*, and *SellerType*. We restricted matches to vehicles sold in the same week. We coarsened *SellerPctArbitraged* into bins of width 0.1[11] and *Valuation* and *Mileage* into bins of width 1,000, and we only allowed matches between vehicles in the same bins.[12] Because there were essentially no potential control vehicles available after 2007 (because the webcast channel was almost fully deployed by then; see Figure 2), we restricted the analysis to observations between 2003 and 2007. The matching procedure yielded 19,095 matched strata that each contained at least one treated vehicle and at least one control vehicle.

To ensure that the procedure resulted in comparable matches, we examined the balance between the treated and control observations as follows. First, we set *DayOfWeek* = 1...7 based on which day of the week a vehicle was sold, with Monday = 1. Second, we calculated the means of *Valuation*, *Mileage*, *SellerPctArbitraged*, and *DayOfWeek* for the treated and control vehicles in each of the 19,095 strata in the matched sample. We then used a t-test to examine whether these strata means differed significantly between the treated and control groups. As shown in Table 4, only *SellerPctArbitraged* showed a significant difference between the two groups after matching, and this difference was of minimal practical significance (recall that matches are exact for all other variables). Table 4 also compares the variables across groups before matching. Overall, we believe that our matches are precise enough to satisfy the unconfoundedness condition (aka, selection on observables) for valid estimation (Imbens 2004). However, it is possible that *unobserved* differences could make the control vehicles inappropriate counterfactuals for the treated vehicles. For this to be a problem, the following conditions would have to hold. First, there would have to be unobserved vehicle characteristics (not captured in our matching procedure) that are correlated with a vehicle being arbitraged. Second, sellers and managers at the facilities—who collectively determine the lane in which vehicles are offered—would have to know what these characteristics are. Third, sellers and facility managers would have to consistently offer vehicles with these characteristics in nonwebcast enabled lanes, while offering the other vehicles in enabled lanes (or vice versa). Although we cannot be sure that these conditions do not hold, they are unlikely. First, it is unlikely that sellers and facility managers would be able to identify the variables—beyond those included in our matching—that consistently predict arbitrage, partly because arbitrage occurs rarely. Also, sellers might want to predict arbitrage if by doing so they could identify "mis"-distributed vehicles and move them to a more advantageous selling location, thereby retaining the arbitrageurs' profits for themselves. However, Overby and Clarke (2012) showed that sellers have little incentive to do this, because the revenues they forgo when they misdistribute a vehicle that is later arbitraged are minimal compared to their total revenues. Last, there may be unobserved variables (e.g., scratches, dents) that influence both whether a vehicle is offered in a webcast enabled lane and its price. Yet if such variables cause a price discount at an arbitrageur's source location, then they will also cause a discount at the destination location, such that potential arbitrage profits would be unaffected. Thus, such variables should not affect the likelihood of arbitrage. We also considered whether buyer heterogeneity across the webcast enabled and nonwebcast enabled lanes might confound our result. We found this to be unlikely, because almost all buyers who purchased in the webcast lanes also purchased in the nonwebcast enabled lanes, and vice versa (see the online appendix for details). We also conducted a sensitivity analysis (Rosenbaum 2002) to assess how influential any unobservables would have to be to alter our conclusion (see below).

Using the matched sample, we fitted the following logistic regression model to test the treatment effect of webcast enablement on whether a vehicle is arbitraged: $\text{logit}(probability(Arbitraged_j = 1)) = \beta_0 + \beta_1 \times WebcastEnabled_j + \varepsilon_j$. We set $Arbitraged_j$ to 1 if vehicle j was arbitraged and 0 otherwise. We set $WebcastEnabled_j$ to 1 if vehicle j was sold in a webcast enabled lane and 0 otherwise. We fitted the model on the matched sample using weighted regression, with the weights provided by the CEM procedure (see Footnote 12). We set $\alpha = 7$ to delineate whether a vehicle was arbitraged in our focal model, and we varied this for robustness. We also fitted the model using two alternative specifications: (a) a rare events logistic regression model, and (b) a linear probability model of the form $Arbitraged_j = \beta_0 + \beta_1 \times WebcastEnabled_j + \varepsilon_j$. Results are virtually identical to those we report. In other unreported analysis, we estimated the model after adding $Valuation_j$, $Mileage_j$, $SellerPctArbitraged_j$, $SaleDate_j$, and indicator variables for each $FacilityID_j$ as explanatory variables. These variables are already accounted for via the matching procedure, and their inclusion has no substantive effect on the $WebcastEnabled_j$ coefficient.

Table 5. Treatment Effects of the Vehicle Being Purchased on a Webcast Enabled Lane (Panel A) and in the Standalone Electronic Channel (Panel B) on Whether the Vehicle Is Later Arbitraged

Panel A. Webcast channel	
$WebcastEnabled_j$ (β_1)	−0.351 (0.137)***
Intercept (β_0)	−5.494 (0.096)***
n	62,837
Log likelihood; $\chi^2_{(1)}$	−1,425; 6.55***
Panel B. Standalone electronic market	
$StandaloneElectronicMarket_j$ (β_1)	0.644 (0.102)***
Intercept (β_0)	−5.596 (0.075)***
n	83,611
Log likelihood; $\chi^2_{(1)}$	−2,416; 39.61***

Notes. The dependent variable is the probability that the vehicle is later arbitraged. Model estimated via logistic regression. Standard errors are in parentheses.
$^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

Panel A of Table 5 shows the results. The treatment effect of webcast enablement is captured by β_1. The effect is negative and significant; treated vehicles (i.e., those sold in webcast enabled lanes) were 29%–39% less likely to be arbitraged than were control vehicles, depending on the value of α.[13] This indicates that the webcast channel reduced the number of spatial arbitrage transactions. We also tested the effect of the webcast channel using McNemar's (1947) test. Because this test requires matched pairs, we restricted the matched sample to only those strata that contained one treated and one control vehicle ($n = 11{,}873$ strata). For each of the matched pairs, one of four outcomes is possible: (a) both the control and the treated vehicle were arbitraged ($n = 1$ in our case), (b) only the control vehicle was arbitraged ($n = 77$), (c) only the treated vehicle was arbitraged ($n = 52$), and (d) neither vehicle was arbitraged ($n = 11{,}743$). McNemar's (1947) test examines whether the probabilities for the (b) and (c) outcomes differ. As above, treated vehicles were significantly less likely to be arbitraged than were control vehicles ($\chi^2_{(1)} = 4.84$; $p < 0.05$). The percentage decrease between (b) and (c) is 32%, similar to the effect size from the main model.

We examined how sensitive our results might be to the possibility that unobserved variables cause vehicles that are likely to be arbitraged to be offered in nonwebcast enabled lanes. If that were the case, then the effect we observe might be attributable to these

Table 6. Descriptive Statistics for Treated and Control Observations for Testing the Effect of the Standalone Electronic Market, Before and After Matching

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

Note. Observations are from 2005 to 2010.

unobservables instead of the webcast channel. We ran a sensitivity analysis based on McNemar's (1947) test to see how influential these unobservables would have to be to alter our conclusion (see Rosenbaum 2002, Section 4.3.2 for details on this procedure). To attribute the higher rate of arbitrage in nonwebcast enabled lanes to unobservables, the unobservables would need to (a) be a near perfect predictor of arbitrage, and (b) produce a 17% increase in the odds of a vehicle being offered in a nonwebcast enabled lane (i.e., $\Gamma = 1.17$ in sensitivity analysis notation). Although there is no consensus about the appropriate size for Γ in social science research, $\Gamma = 1.2$ is average (Sen 2014).[14]

5.2.2. Testing the Effect of the Standalone Electronic Market.

To examine the effect of the standalone electronic market on spatial arbitrage, we used a similar matching procedure as above, with one major change. We considered vehicles sold in the standalone electronic market to be potential treated vehicles, with vehicles sold in the physical market—regardless of whether they were sold in a webcast enabled lane—as potential control vehicles. We ran the matching procedure using the data from January 1, 2005, to December 31, 2010, given minimal transaction volume in the standalone electronic market prior to 2005. The matching procedure yielded 26,937 strata consisting of 83,611 vehicles: 29,612 treated vehicles and 53,999 control vehicles. The balance between treated and control observations was good except for *DayOfWeek* (Table 6). For robustness, we assessed whether the imbalance for *DayOfWeek* affected our results by exact matching on observations with the same *DayOfWeek*. These results are consistent with those we report below. Our regression specifications were identical to those for analyzing the effect of the webcast channel, except we replaced $WebcastEnabled_j$ with $StandaloneElectronicMarket_j$.

Results of the logistic regression model appear in panel B of Table 5; results of the other models are virtually identical. The β_1 coefficient is positive and significant. Vehicles purchased in the standalone electronic market are 55%–117% more likely to be arbitraged than vehicles in the physical market, depending on the value of α. We also tested the effect of the standalone electronic market using McNemar's (1947) test, as above. In this test, the "both arbitraged" outcome has $n = 2$ observations, the "control arbitraged, treated not" outcome has $n = 64$, the "treated arbitraged, control not" outcome has $n = 117$, and the "both not" outcome has $n = 15{,}906$. Treated vehicles were significantly more likely to be arbitraged than were control vehicles ($\chi^2_{(1)} = 15.52$; $p < 0.01$); the percentage increase is approximately 83%.

Sellers decide whether to offer vehicles in the standalone electronic market or the physical market, and they also decide what prices to accept. If unobserved factors influence these decisions and whether a vehicle is likely to be arbitraged, then our result could be biased. One possibility is that sellers accept below-market prices for vehicles in the standalone electronic market (i.e., they "dump" vehicles for less than they are worth), which would create arbitrage opportunities. This does not appear to be the case (on average), because the *Price/Valuation* ratio is significantly higher for vehicles sold in the standalone electronic market than in the physical market (101% to 99%; $p < 0.01$; see Overby and Mitra (2014) for more on this discrepancy). Another possibility is that sellers offer in the standalone electronic market vehicles with unobserved vehicle characteristics that are correlated with arbitrage. For this to be true, sellers would have to (a) know which unobservables are correlated with arbitrage, (b) identify vehicles that have these unobservables, and (c) offer these vehicles consistently on the standalone electronic market and not in the physical market. For the reasons described in Section 5.2.1, we believe this to be unlikely. In addition, we assessed how sensitive our results are to the possibility that unobserved variables cause vehicles that are inherently likely to be arbitraged to be offered in the standalone electronic market (as above). To attribute the higher rate of arbitrage in the standalone electronic market to these unobservables, they would need to predict arbitrage almost perfectly, and they would need to produce more than a 48% increase in the odds of a vehicle being offered in the standalone electronic market ($\Gamma = 1.49$). Also, see the online appendix for a discussion of why buyer heterogeneity is unlikely to confound our conclusion.

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

5.2.3. Analysis of the Mechanisms Behind the Effects of the Two Electronic Channels. Our theory suggests that the expanded reach that electronic commerce provides should reduce arbitrage opportunities by helping remotely located buyers—who might otherwise be potential "downstream" customers for arbitrageurs—purchase vehicles directly from source locations. Our theory also suggests that expanded reach should help arbitrageurs identify and exploit otherwise hidden arbitrage opportunities. It also suggests that arbitrageurs should leverage transaction immediacy to exploit arbitrage opportunities originating in the standalone electronic market. Whether these mechanisms increase or decrease the number of arbitrage transactions depends on whether the opportunity reduction effect outpaces the opportunity exploitation effect.

To explore this for the webcast channel, we first confirmed that the webcast channel led to expanded buyer reach. We used the matched sample and the regression specifications from Section 5.2.1 to test the treatment effect of a vehicle being offered in a webcast enabled lane on the likelihood of its being purchased by a remotely located buyer (*RemoteBuyer*). We set *RemoteBuyer* = 1 if the distance between the buyer and the facility at which the vehicle was located (i.e., *BuyerDistance*; see Table 3) was at least one standard deviation above the mean (two standard deviations for robustness). Vehicles purchased from webcast enabled lanes were 15%–35% more likely to be purchased by a remote buyer ($p < 0.01$), depending on the measure of "remote."[15] We also tested the treatment effect of the webcast channel on vehicle price (*Price*), finding a 0.5% increase ($p < 0.01$), perhaps due to an increase in the number of buyers bidding on the vehicles (although our data do not report the number of bidders). The *RemoteBuyer* result could reflect both the opportunity reduction effect (due to regular buyers purchasing at remote locations) and the opportunity exploitation effect (due to arbitrageurs finding vehicles in remote locations that can be profitably arbitraged). Yet the *Price* result should pertain only to opportunity reduction because it increases the arbitrageurs' cost of sourcing vehicles. The opportunity reduction effect seems to outpace the opportunity exploitation effect, generating the negative effect of the webcast channel on spatial arbitrage transactions.

We used an analogous procedure to test the effects of the standalone electronic market on *RemoteBuyer* and *Price*. Results are similar; vehicles purchased in the standalone electronic market were 72%–83% ($p < 0.01$) more likely to be purchased by a remote buyer (depending on the measure of "remote") and had 1.1% higher prices ($p < 0.01$). A key difference between the webcast channel and the standalone electronic market is that the latter provides transaction immediacy, which should help arbitrageurs identify and exploit arbitrage opportunities (see Section 4.1). If this is happening, then we should see arbitrageurs purchasing undervalued vehicles in the standalone electronic market very soon after they are listed there. To explore this, we obtained supplemental data that contained additional variables for 55% ($n = 337{,}295$) of the transactions that occurred in the standalone electronic market between 2007 and 2010.[16] We used two variables in particular: the date the vehicle was listed on the standalone electronic market (*ListingDate*) and whether a vehicle was purchased via auction (either before or after the reserve price was met) or via Buy Now (*TransactionType*). Using the supplemental data, we calculated *DaysToSale* as the number of days between a vehicle's *ListingDate* and *SaleDate*. We also calculated *PriceValRatio* as the ratio of a vehicle's *Price* to its *Valuation*. We computed the means of *PriceValRatio* and *DaysToSale*, in aggregate and by *TransactionType*, for all transactions and for only those that were spatially arbitraged ($n = 1{,}967$). Results appear in Table 7.

Consistent with our theory, arbitrageurs used the standalone electronic market to purchase undervalued vehicles more successfully than did regular buyers: the mean *PriceValRatio* paid by arbitrageurs was 95%, whereas the overall mean was 101% (the difference is significant at $p < 0.01$). Also, arbitrageurs purchased vehicles very soon after they were listed: arbitrageurs waited only 0.81 days (on average), which is less than half the overall average of 1.76 days (the difference is significant at $p < 0.01$). The time between listing and purchase for arbitrageurs was particularly short for Buy Now transactions and for auction transactions that occurred before the reserve price was met. This supports our theory that arbitrageurs use the standalone electronic market to scan for vehicles with undervalued Buy Now prices, leveraging the market's transaction immediacy to purchase them quickly. It also suggests that arbitrageurs are quick to register "low-ball" bids for vehicles that—if accepted—lead to profitable arbitrage opportunities. Although many of

Table 7. Statistics for *PriceValRatio* and *DaysToSale* in the Supplemental Data

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

these bids are likely beaten, some sellers accept them, perhaps because they represent a quick way to sell a vehicle. Neither of these behaviors is available to an arbitrageur when sourcing vehicles in the physical market, regardless of whether he is participating via webcast or physically. We examined whether these undervalued vehicles purchased via the standalone electronic market were particularly likely to be arbitraged. As shown in Table 8, undervalued vehicles (i.e., those for which *PriceValRatio* < 1) were more

Table 8. Percentage of Vehicles Arbitraged from 2007 to 2010, By *TransactionType* and Degree to Which the Vehicle Was Undervalued Relative to Its Valuation

	Standalone electronic market transactions (%)			
	Buy now	Auction: Above reserve	Auction: Below reserve	All transactions (%)
All vehicles	0.54	0.58	0.97	0.58
Undervalued vehicles				
(a) Vehicles with *PriceValRatio* < 1	1.04	1.01	1.30	0.91
(b) Vehicles with *PriceValRatio* < 0.99	1.18	1.06	1.32	0.94
(c) Vehicles with *PriceValRatio* < 0.95	1.68	1.48	1.75	1.04

Notes. Statistics for standalone electronic market transactions are from the supplemental data described in the text. Statistics are based on the 2007 to 2010 time period to correspond to the supplemental data. All percentages for "All Vehicles" are different ($p < 0.10$) except for those for (1) "All Transactions" and "Auction: Above Reserve" and (2) "Buy Now" and "Auction: Above Reserve". All percentages for "Undervalued Vehicles" are different ($p < 0.10$) except for those for (1) "Buy Now" and "Auction: Above Reserve" in row a, (2) "Buy Now" and "Auction: Below Reserve" in rows (b) and (c), and (3) "Auction: Above Reserve" and "Auction: Below Reserve" in row (c).

likely to be arbitraged than were overvalued vehicles (as one would expect), with this being particularly true for vehicles purchased via Buy Now and for below their auction reserve price in the standalone electronic market.[17]

We also estimated the profit for each arbitrage transaction using the formula noted in Section 5.1. The average arbitrage profit is $781 (st. dev. $1,030) when the vehicle was sourced from the standalone electronic market and $672 (st. dev. $710) when the vehicle was sourced from the physical market via either the traditional physical channel or the webcast channel (this difference is significant at $p < 0.01$). Also, the percentage of arbitrage transactions that were profitable is higher for vehicles sourced in the standalone electronic market (91.6% versus 88.6%; $p < 0.01$). The overall pattern of results suggests that the arbitrage opportunity exploitation effect is stronger in the standalone electronic market than in the webcast channel because the former supports transaction immediacy. As a result, the opportunity exploitation effect seems to outpace the opportunity reduction effect in the standalone electronic market, generating the positive effect of the standalone electronic market on spatial arbitrage transactions.

5.3. Testing H3 and H4: Factors That Affect Arbitrageur Behavior of Where to Source Products

We used a discrete choice model to study how arbitrageurs choose where to source vehicles. Fitting a choice model requires the researcher to define the set of alternatives available to the decision maker (referred to as the *choice set*) and to specify a utility function for each alternative (Train 2009). We observe the facility at which arbitrageur i on day t sourced a vehicle(s) that he later arbitraged; this is the "chosen" alternative in each choice set. We defined the "nonchosen" alternatives in the choice set as those facilities other than the chosen facility: (a) that were open on day t, and (b) at which arbitrageur i made a purchase during the sample period. We modeled the utility of each facility k to arbitrageur i at time t as $U_{ikt} = \beta_{0,k} + \beta_1 \times PctOfferedWebcast_{kt} + \beta_2 \times Distance_{ik} + \beta_3 \times Distance_{ik} \times NearbyFacilities_k + \beta_4 \times Supply_{kt} + \beta_5 \times Supply^2_{kt} + \beta_6 \times PctSold_{k(t-30)} + \beta_7 \times PctSoldLowPrice_{k(t-30)} + \beta_8 \times GeoPriceDispersion_{k(t-30)} + \varepsilon_{ikt}$. We describe the explanatory variables in the utility function in Table 9.

We included $PctSoldLowPrice_{k(t-30)}$ and $GeoPriceDispersion_{k(t-30)}$ to test arbitrageur preferences for facilities at which a high percentage of vehicles recently sold for low prices and at which prices for the vehicles recently offered varied widely across facilities (H3A and H3B). We included $Distance_{ik}$ to test arbitrageur preferences for facilities close to them (H4A). We tested H4B (that arbitrageurs prefer facilities that are relatively difficult for other traders to access) in two ways.

Table 9. Variables Used in the Discrete Choice Model of Where Arbitrageurs Source Vehicles

Variable	Description	Mean (st. dev.)
$PctOffered\ Webcast_{kt}$	The percentage of vehicles at facility k that were offered in webcast enabled lanes on day t.	0.62 (0.41)
$Distance_{ik}$	The distance in miles between the zip codes of arbitrageur i and facility k.	535.23 (528.18)
$NearbyFacilities_{k}^{a}$	The number of facilities within 350 miles of facility k.	20 (8)
$Supply_{kt}$	The number of vehicles offered at facility k on day t.	783.40 (765.07)
$PctSold_{k(t-30)}$	The percentage of vehicles offered at facility k in the 30 days prior to day t that were sold.	0.61 (0.12)
$PctSoldLow\ Price_{k(t-30)}^{b}$	The percentage of vehicles sold at facility k in the 30 days prior to day t that sold for less than 90% of their *Valuation*.	0.20 (0.10)
$GeoPriceDispersion_{k(t-30)}$	The average geographic price dispersion of the vehicles offered at facility k in the 30 days prior to day t. We measured this as follows. First, for each vehicle offered at facility k on day t, we created a list of all facilities at which vehicles of the same year/model were purchased in the 30 days prior to day t. Second, we calculated the average price for those vehicles at each facility. Third, we took the standard deviation of these average prices across facilities. This gave us a measure of how much the price of each vehicle offered at facility k on day t varied across facilities. Fourth, we averaged these measures for all vehicles offered at facility k on day t.	1,537.77 (242.00)

[a]We also ran the model using 100, 200, and 700 mile thresholds to define $NearbyFacilities_k$. Results are similar.
[b]We also ran the model using 80% and 85% thresholds to define $PctSoldLowPrice_{k(t-30)}$. Results are similar.

First, we interacted $Distance_{ik}$ with $NearbyFacilities_k$, reasoning that facilities located near other facilities are relatively easy for other traders to access, given that the location of the facilities closely matches the population density of the United States.[18] Second, we included $PctOfferedWebcast_{kt}$, reasoning that the more the vehicles at a facility are available via webcast, the more accessible (electronically) that facility is to traders. Furthermore, including $PctOfferedWebcast_{kt}$ allowed us to test whether arbitrageurs prefer sourcing vehicles at facilities at which the webcast channel has been deployed only minimally or not at all, which one might expect given the negative effect of the webcast channel on spatial arbitrage transactions shown in Section 5.2.1. We included $Supply_{kt}$, $Supply_{kt}^2$, and $PctSold_{k(t-30)}$ to account for the role of facility size and liquidity. We used 30-day lagged variables (noted by the $t-30$ subscript) for $PctSold_{k(t-30)}$, $PctSoldLowPrice_{k(t-30)}$, and $GeoPriceDispersion_{k(t-30)}$ because the contemporaneous values of these variables are unknown to arbitrageurs when they choose the facility at which to purchase. We used contemporaneous variables for the other variables because arbitrageurs already know them (e.g., $Distance_{ik}$) or can calculate them based on the presale list posted in advance on the intermediary's website. We included alternative-specific constants to capture the latent utility (i.e., the fixed effect) of each facility k (represented as the $\beta_{0,k}$ term) and fitted the model using the multinomial logit specification. Results appear in Table 10.

H3A is supported; the coefficient for $PctSoldLowPrice_{k(t-30)}$ is positive and significant. We used the model estimates to simulate the size (i.e., practical significance) of this effect as follows. We simulated the percentage change in the number of times arbitrageur i chose facility k when $PctSoldLowPrice_{k(t-30)}$ = 10% versus 30% (i.e., one standard deviation below and above the mean). We did this for each facility. On average, this increased the probability of choosing a facility by approximately 20%. H3B is not supported; the coefficient for $GeoPriceDispersion_{k(t-30)}$ is positive (as posited) but insignificant. We explore this further in the online appendix. H4A is supported; the coefficient for $Distance_{ik}$ is negative and significant. We simulated the size of the $Distance_{ik}$ effect by estimating the percentage change in the number of times an

Table 10. Results of the Discrete Choice Model of Where Arbitrageurs Source Vehicles

	Coefficient (std. error)
$PctOfferedWebcast_{kt}$ (β_1)	−0.3420 (0.0225)***
$Distance_{ik}$ (β_2)	−0.5025 (0.0322)***
$Distance_{ik} \times NearbyFacilities_k$ (β_3)	−0.0383 (0.0017)***
$Supply_{kt}$ (β_4)	3.5414 (0.0217)***
$Supply_{kt}^2$ (β_5)	−0.5477 (0.0054)***
$PctSold_{k(t-30)}$ (β_6)	−2.1817 (0.0606)***
$PctSoldLowPrice_{k(t-30)}$ (β_7)	1.0042 (0.1072)***
$GeoPriceDispersion_{k(t-30)}$ (β_8)	0.0041 (0.1030)
Alternative specific constants	Included
Number of alternatives per choice set (min, mean, max)	2, 5.9, 48
n (total number of choices)	698,057
Log likelihood	−115,480

Notes. The number of days used to delineate spatial arbitrage (α) is set to seven in this analysis. Results are consistent for $\alpha = 5$, $\alpha = 14$, and $\alpha = 28$. Standard errors are in parentheses. To ensure that coefficients are of similar magnitude for reporting purposes we divided $Supply_{kt}$, $Distance_{ik}$, and $GeoPriceDispersion_{k(t-30)}$ by 1,000.
$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

arbitrageur chose facility k when $Distance_{ik} = 250$ versus when $Distance_{ik} = 750$, with $NearbyFacilities_k$ set at its mean, which is 20. The increased distance reduced the probability of choosing a facility by 21%. H4 is also supported; the interaction between $Distance_{ik}$ and $NearbyFacilities_k$ is negative and significant. This shows that if two facilities are equidistant, the arbitrageur will prefer the one that is more isolated. To examine the effect of $NearbyFacilities_k$, we set $Distance_{ik}$ at its mean and ran the simulations with $NearbyFacilities_k$ set to 12 and 28. This increased density (and greater accessibility) reduced the probability of choosing the facility by 14%. The negative and significant coefficient for $PctOfferedWebcast_{kt}$ also provides support for H4B, along with corroborating our earlier results about the effect of the webcast channel. To examine the size of this effect, we ran the simulations with $PctOfferedWebcast_{kt} = 24\%$ and $PctOfferedWebcast_{kt} = 71\%$, which are the mean values for this variable in 2003 and 2004 (and represent close to a one standard deviation increase). This reduced the probability that an arbitrageur would choose the facility by 13%. The supply of vehicles (i.e., $Supply_{kt}$) has a positive effect. This may be because a large supply increases an arbitrageur's chance of finding undervalued vehicles. This relationship is concave (i.e., $Supply^2_{kt}$ has a negative effect), but the inflection point does not occur until $Supply_{kt} = 2{,}532$, which is more than two standard deviations above the mean (note that we scaled $Supply_{kt}$ by dividing by 1,000; see Table 10). Arbitrageurs also prefer to source at facilities at which recent sales percentages have been low ($PctSold_{k(t-30)}$ is negative), i.e., those that have been relatively illiquid.

6. Conclusion

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

6.1. Contributions and Summary of Findings

We make three main contributions. Our first contribution is that we measure market efficiency via the existence and exploitation of spatial arbitrage opportunities rather than via price dispersion. The spatial arbitrage measure has several advantages over price dispersion for markets in which trading is geographically distributed. First, it inherently accounts for the transaction costs of moving products between locations. If these transaction costs are lower than the price difference between locations (i.e., if supply and demand are inefficiently distributed), then arbitrage will occur. If not, it will not. Second, the spatial arbitrage measure accounts for unobserved product heterogeneity because the same product is traded at both the source and destination locations. Unobserved product heterogeneity can confound the price dispersion measure if the products whose prices are being compared across locations differ because of unobserved quality differences. Third, the spatial arbitrage measure is based on the micro-level behavior of the agents—the arbitrageurs—who are most aware of whether buyers and sellers are matching efficiently across geography. Thus, arbitrageur behavior provides a direct window into a market's level of efficiency. A potential drawback to the spatial arbitrage measure is that arbitrage transactions are difficult to observe. Using spatial arbitrage as the measure requires unique (and consistent) identifiers for individual products, traders, and market locations. Although such data are relatively elusive, they are becoming more common as trading activity increasingly moves online and item-level tracking becomes more widely adopted (e.g., see the Livestock Identification System in Australia; http://www.dpi.nsw.gov.au/agriculture/livestock/nlis.

Our second contribution is that we study two distinct forms of electronic commerce. This helps us understand the theoretical mechanisms through which electronic commerce affects market efficiency, and it helps us develop a more nuanced and complete understanding of how electronic commerce affects market efficiency than has been previously shown empirically (to our knowledge). Both forms of electronic commerce—the webcast channel and the standalone electronic market—provide expanded reach. This should reduce spatial arbitrage opportunities by helping regular buyers purchase directly from source locations (thereby disintermediating arbitrageurs) while simultaneously improving arbitrageurs' ability to find and exploit arbitrage opportunities that he might otherwise miss. Yet only the standalone electronic market provides transaction immediacy, which should further enhance arbitrageurs' ability to find and exploit arbitrage opportunities before they dissipate. Our results reveal this nuanced effect: electronic commerce reduces arbitrage opportunities, but it improves arbitrageurs' ability to find and exploit those that remain. This shows that electronic commerce improves market efficiency in two ways. First, electronic commerce helps buyers and sellers trade across distance, thereby balancing supply/demand across geographic locations (as has been shown in prior research). Second, electronic commerce improves arbitrageurs' ability to identify and exploit remaining supply/demand imbalances, which rebalances supply and demand and restores efficiency (which has not been shown in prior research). Only one of the forms of electronic commerce in our study supports transaction immediacy, and we used this variation to disentangle how electronic commerce improves arbitrageurs' ability to exploit (and thereby correct) supply/demand imbalances. If both forms supported transaction immediacy (hypothetically), then efficiency would likely have been further improved.

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

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

6.2. Limitations and Future Research

A limitation of our analysis is that it is specific to the wholesale used vehicle industry. However, well-functioning automotive markets are important in their own right, given the surprisingly large impact that the automotive industry has on the overall U.S. economy.[19] This importance is reflected in several academic studies focused on the industry (e.g., Dimoka et al. 2012). Testing whether the results hold for other industries represents an opportunity for future research. Another limitation of our analysis is that many of the results are based on matching estimation in a quasi-natural experimental setting. We have matched on many important variables, including market facility, sale date, vehicle year, make, model, mileage, valuation, seller type, and the probability that a seller's vehicles are arbitraged. However, we cannot rule out the possibility that unobserved variables might confound our conclusions, although we have conducted sensitivity analyses to assess how influential these unobservables would need to be. Another limitation is that although we are able to observe spatial arbitrage transactions with a high level of precision, the precision is imperfect. It is possible that we have misclassified spatial arbitrage instances, although our results are robust to different measures. Last, space and scope limitations preclude us from examining other research questions, such as how arbitrageurs choose between the physical market and the standalone electronic market when sourcing vehicles for arbitrage and/or when selling vehicles to complete arbitrage transactions. This represents an opportunity for future research.

Acknowledgments

The authors would like to thank the senior editor, associate editor, and three anonymous reviewers for their help and guidance on this paper. The authors also appreciate the comments and suggestions provided by seminar participants at the Georgia Institute of Technology, the University of Minnesota, the University of Florida, Carnegie Mellon University, Florida International University, Temple University, the University of Georgia, the University of Cincinnati, the University of Memphis, the 2013 Conference on Information Systems and Technology, and the 2015 Statistical Challenges in E-Commerce Research Symposium.

Endnotes

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

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

[3] We include the word "typically" because it is possible for an electronic commerce system to only allow transactions to be conducted

at certain times. One of the electronic channels that we study operates this way.

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

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

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

[7] This approach assumes that moving a matched vehicle from the source facility to the destination facility would not change the estimated price at the destination facility. This is questionable, because the additional supply at the destination facility would likely lower prices. This will cause our estimates of the number of arbitrage opportunities to be biased upward. However, this bias will be consistent across all eight years of our sample. Given this consistency, the year-over-year decline shown in Figure 3 can still be interpreted as a decline in the number of arbitrage opportunities. Any other form of mismeasurement that exists across years will also not affect our conclusion.

[8] To verify that the declines shown in panels A and B of Figure 3 are not artifacts of a decline in overall transaction volume, we also calculated the time series for the percentages of arbitrage opportunities and arbitrage transactions relative to total market transactions. These time series mirrored those shown in panels A and B of Figure 3 ($\rho = 0.99$ and $\rho = 0.98$).

[9] The reduction in arbitrage transactions (panel B of Figure 3) suggests that the (negative) opportunity reduction effect of electronic commerce dominates the (positive) opportunity exploitation effect overall.

[10] As support for this, we calculated the following for the first quarter of 2003: (a) the number of vehicles of each *Year/Make/Model* sold in webcast enabled lanes ($\mu = 5.4$, $\sigma = 42.1$), (b) the number of vehicles of each *Year/Make/Model* sold overall ($\mu = 156.4$, $\sigma = 601.8$), and (c) the number of facilities at which vehicles of each *Year/Make/Model* were sold ($\mu = 22.7$, $\sigma = 23.4$). We regressed (a) on (b) and (c). The coefficient for (c) was −0.09 ($t = -4.90$), such that a one standard deviation increase in the number of facilities at which a vehicle was sold was associated with a 39% decrease in the number of vehicles sold in webcast enabled lanes. We are also aware of the procedure by which vehicles were assigned to webcast enabled lanes because one of the authors consulted with the intermediary on the initial implementation of the webcast channel.

[11] Because $SellerPctArbitraged \leq 0.004$ for over half of the observations, we further coarsened the 0 to 0.1 bin into the following bins: 0 to 0.0003, 0.0003 to 0.0006, 0.0006 to 0.0009, . . . , 0.0033 to 0.0036, 0.0036 to 0.004, and 0.004 to 0.1. We also used a single bin for values between 0.5 and 1, given the rarity of observations with these values.

[12] CEM temporarily coarsens each chosen variable into bins and exact matches on those bins, yielding a matched sample of control and treated observations. CEM then restores the original (noncoarsened) values of the variables for analysis. The CEM matching procedure may match an uneven number of control observations to treated observations. To account for this, CEM generates weights. Using these weights in an analysis procedure (such as regression) generates the sample average treatment effect. See Iacus et al. (2012) for details.

[13] We also used $\alpha = 5$, 14, and 28 days. We obtained the 29%–39% estimates by exponentiating the β_1 coefficients and by analyzing the results of the linear probability model (we divided β_1 by β_0 from the linear probability model).

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

[15] This corroborates a similar finding in Overby and Forman (2015).

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

[17] When all vehicles (undervalued and overvalued) are considered, the percentage of Buy Now transactions that are arbitraged is slightly below that for all transactions. Analysis by Overby and Mitra (2014) provides a likely explanation. They conclude that Buy Now is used frequently by regular buyers who pay over wholesale market value to purchase specific vehicles for retail customers, with whom they have likely already negotiated a retail price. This is consistent with the mean *PriceValRatio* for Buy Now transactions being 1.02 (see Table 7). Our results show that when Buy Now is used for an undervalued vehicle, that vehicle has a (relatively) high likelihood of being arbitraged.

[18] We did not include $NearbyFacilities_k$ as a standalone variable because it is a constant for each facility k and cannot be estimated separately from each facility's alternative-specific constant, which is represented by $\beta_{0,k}$.

[19] In 2012, sales at automobile dealerships represented approximately 15% of total retail sales in the United States, and dealership payroll represented approximately 12% of total retail payroll (National Automobile Dealers Association; https://www.nada.org/nadadata/).

References

Aker JC (2010) Information from markets near and far: Mobile phones and agricultural markets in Niger. *Amer. Econom. J.: Appl. Econom.* 2(3):46–59.

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

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

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

Barrett C (2008) Spatial market integration. Durlauf S, Blume L, eds. *The New Palgrave Dictionary of Economics* (Palgrave Macmillan, London), 6297–6300.

Baye MR, Morgan J, Scholten P (2006) Information, search, and price dispersion. Hendershott T, ed. *Handbook on Economics and Information Systems* (Elsevier, Amsterdam), 323–376.

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

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

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

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

Dimoka A, Hong Y, Pavlou PA (2012) On product uncertainty in online markets: Theory and evidence. *MIS Quart.* 36(2): 395–426.

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

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

Iacus SM, King G, Porro G (2012) Causal inference without balance checking: Coarsened exact matching. *Political Anal.* 20(1): 1–24.

Imbens G (2004) Nonparametric estimation of average treatment effects under exogeneity: A review. *Rev Econom. Statist.* 86(1): 4–29.

Jensen R (2007) The digital provide: Information (technology), market performance, and welfare in the south Indian fisheries sector. *Quart. J. Econom.* 122(3):897–924.

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

McNemar Q (1947) Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12(2):153–157.

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

Overby E, Mitra S (2014) Physical and electronic wholesale markets: An empirical analysis of product sorting and market function. *J. Management Inform. Systems* 31(2):11–45.

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

Parker C, Ramdas K, Savva N (2016) Is IT enough? Evidence from a natural experiment in India's agriculture markets. *Management Sci.* 62(9):2481–2503.

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

Persson KG (2008) The law of one price. Whaples R, ed. *EH.Net Encyclopedia of Economic and Business History* (Economic History Association, Tucson, AZ).

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

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

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

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

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

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