

The Effect of Electronic Commerce on Market Integration and Spatial Arbitrage

The bulk of this paper was done by a student.

Abstract: Markets that are well-integrated spatially are important to the stability and success of the economy because they balance supply and demand and create incentives for innovation. Electronic commerce has the potential to improve market integration by providing visibility into supply/demand conditions in different geographic regions and facilitating trade across geographic distance. We study the effect of electronic commerce on market integration as measured by the prevalence of spatial arbitrage, which has several advantages over the price-based measures used in prior research. We also study the effect of two distinct forms of electronic commerce, which helps us better identify the linkages between electronic commerce, spatial arbitrage, and market integration. Based on an analysis of over 40 million transactions in the wholesale used vehicle industry from 2003 to 2010, we conclude that electronic commerce improved market integration.

1.0 Introduction, Motivation, and Literature Review

Spatially integrated markets – i.e., those in which buyers and sellers can easily trade across geographic space with minimal friction caused by informational, infrastructural, or regulatory barriers – are important for several reasons. If markets are well-integrated spatially, then goods can move easily from geographic regions with abundant supply to regions of scarcity. If markets are *not* well-integrated, then prices for scarce goods may rise dramatically in regions experiencing high demand, creating hardship in those regions. Similarly, if a seller is unable to reach buyers in geographically remote regions, then a technological innovation that boosts her production may simply create excess supply in her local region, leading to price drops that may negate potential gains and discourage innovation. Given the importance of well-integrated markets, developing methods for measuring markets' spatial integration and assessing what factors foster or inhibit it are important social science research topics (Baulch 1997).

Electronic commerce has the potential to foster market integration by increasing the visibility of price information and by facilitating trade across regions, thereby helping buyers and sellers distribute demand and supply efficiently. However, it has been challenging to determine the effect of electronic commerce (or of any other IT-based innovation) on market integration because of difficulty measuring the degree to which a market is integrated. Scholars have measured market integration using price-based measures such as geographic price dispersion (Aker 2010; Jensen 2007; Overby and Forman 2011) and the co-movement of prices at different locations over time (Alexander and Wyeth 1994; McNew 1996; Ravallion 1986). Although useful, these methods have limitations because: a) they may not properly consider the transaction costs of trading across geographic distance, and b) they may be inaccurate due to unobserved product quality differences across locations (Barrett 2005).

To overcome these limitations, we propose a different measure of market integration: the prevalence of spatial arbitrage. *Spatial arbitrage* occurs when a trader (whom we refer to as the *arbitrageur*) exploits a price discrepancy between two locations by purchasing a product at the low-price location and reselling it at the high-price location.¹ Spatial arbitrage should be difficult in well-integrated markets, because buyers (sellers) can shift demand (supply) to low-priced (high-priced) locations directly, eliminating the need for arbitrageurs to serve as “middlemen”. By contrast, if markets are not well-integrated, then spatial arbitrage opportunities will abound. The prevalence of spatial arbitrage has long been considered a powerful measure of a market’s integration for several reasons (Baulch 1997). First, it is based on the micro-level behavior of the traders (the arbitrageurs) who are most aware of supply/demand imbalances that reflect a lack of market integration. Second, it is immune to potential product quality differences across locations, because the same product is traded in both locations. Third, it implicitly accounts for the transaction costs of trading across geography, because arbitrageurs must pay these costs to engage in arbitrage. Despite its merits, spatial arbitrage has been an elusive measure of market integration because of difficulty tracking the flow of the same product across geography (Minten and Barrett 2008). We overcome this by studying spatial arbitrage in the United States wholesale used vehicle industry, where we can track the geographic flow of each vehicle based on its unique Vehicle Identification Number (*VIN*). This allows us to observe instances in which a trader i engages in spatial arbitrage by buying a vehicle at a “source” location j for price p and then quickly reselling the same vehicle at a “destination” location k for a higher price p' .

We study the following research question: *How does electronic commerce affect market integration as measured by the prevalence of spatial arbitrage?* Although using spatial arbitrage

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.” This is because arbitrage, as we define it, requires both a purchase and a sale.

as a measure of market integration has many advantages, the measure must be used carefully. This is because electronic commerce might decrease the prevalence of spatial arbitrage by helping buyers and sellers trade with each other across geographic space, thereby eliminating the arbitrageur middleman. In this case, a decrease in spatial arbitrage would reflect an increase in market integration. However, electronic commerce might also *increase* the prevalence of spatial arbitrage by helping arbitrageurs identify mispriced assets that they can profitably arbitrage. In this case, an *increase* in spatial arbitrage might *not* indicate a *decrease* in market integration, but rather an increased ability to exploit a given level of market *non*-integration. Fortunately, our data and empirical setting help us disentangle these two possibilities to some extent. This is because there were two distinct electronic trading mechanisms implemented in the wholesale used vehicle industry in recent years: a) a webcast channel that allows electronic access to the traditional physical market, and b) a stand-alone electronic market in the form of an online store. Our findings indicate that the webcast channel reduces spatial arbitrage by helping buyers and sellers trade with each other directly, thereby disintermediating arbitrageurs and improving market integration. The standalone electronic market may have a similar effect, but the manner in which it is designed also help arbitrageurs identify mispriced vehicles to arbitrage, and we find that this market actually fosters spatial arbitrage. The webcast effect dominates, such that increasing levels of electronic trading have led to an overall decrease in spatial arbitrage and to a greater level of market integration. Our examination of these two distinct electronic channels is another contribution of our study (in addition to our use of spatial arbitrage to measure market integration), as many prior studies have implicitly considered electronic commerce to be monolithic and have not considered how the design of electronic channels might influence outcomes.^{2,3}

² We have withheld the detailed literature review to adhere to the CIST 15-page limit. This includes

2.0 Empirical Context

The empirical context for our study is the U.S. wholesale used vehicle market. Buyers in this market are used car dealers who use the market as a primary source of inventory for their retail lots. Sellers are of two types: a) used car dealers who use the market to sell vehicles that they cannot (or choose not to) sell retail, and b) institutional sellers such as rental car firms who 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 80 physical market facilities in the U.S. as well as the webcast channel and stand-alone electronic market alluded to above and described below. The data consist of 40,657,724 successful transactions facilitated by the intermediary from Jan 1st 2003 to Dec 31st 2010. There were 34,510,905 unique vehicles traded in the sample, as some vehicles were traded more than once.

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

linkages to the IS literature on price dispersion, including such papers as those by Brynjolfsson and Smith (2000), Chellappa et al. (2011), Clemons et al. (2002), and Ghose and Yao (2011). Although these papers are relevant given their focus on electronic commerce and market efficiency, they typically ignore the location of traders and products, which is critical for examining market integration in a spatial sense. Our focus on location is one of the differentiators of our study. Another differentiator is our analysis of the two distinct electronic channels, as this departs from prior studies that have considered electronic commerce monolithically.

³ Perhaps the closest precedent to our research is the study by Overby and Clarke (2012), who also examined spatial arbitrage in the wholesale used vehicle industry. Overby and Clarke used behavioral economic theory to analyze how the limited attention of sellers affects how they distribute vehicles across locations in the market and the downstream effect this has on the probability that a vehicle is arbitrated. They also developed estimates of the prevalence and profitability of spatial arbitrage. We depart from their work in several ways. First, we focus on how electronic commerce affects market integration as measured by spatial arbitrage. Overby and Clarke (2012) commented on this by presenting the correlation between webcast bidding activity on a vehicle and the probability that it was spatially arbitrated (see their §5.1.2), but this was a secondary result and not their main focus. Further, this correlation was only weakly identified econometrically. By contrast, we leverage natural experiments to estimate the causal effect of electronic commerce on spatial arbitrage. Second, we study not only the webcast channel but also the stand-alone electronic market, which was ignored by Overby and Clarke. We corroborate our findings by examining how electronic commerce affects trade across geographic distance and by studying how arbitrageurs choose where to source vehicles, neither of which was examined by Overby and Clarke.

parking lot for vehicle storage and a warehouse-type building equipped with multiple *lanes*, which are essentially one-way streets. 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 per vehicle; it is common for auctions to be held simultaneously in different lanes at the same facility.

Approximately 10 years ago, the intermediary that 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 physical auctions via a web browser, and it also 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 auctions occurring in the physical market. The webcast channel was implemented in phases as the necessary equipment (e.g., camera, microphone) was installed *in each lane at each facility*.⁴ The second panel of Figure 1 shows how the percentage of vehicles available via webcast increased as the channel was implemented. As discussed below, we leveraged the phased implementation of the webcast channel in a natural experiment to assess the effect of the webcast channel on the probability that a purchased vehicle was later arbitrated.

The intermediary also operates a stand-alone electronic market that functions similarly to eBay. In this market, sellers list their vehicles, and buyers have the option to purchase them for a fixed "Buy Now" price or to bid for them. It is important to note that sellers choose to offer a vehicle either: a) at a physical market facility, where the vehicle is available to buyers present at

⁴ We inferred the webcast implementation date for each lane as the first date a webcast transaction occurred in that lane. To account for a possible lag between webcast implementation and the first transaction, we alternatively set each lane's implementation date as the first webcast transaction date minus 3 weeks. This does not affect our results.

that facility and to buyers accessing that facility via webcast (if enabled), or b) in the stand-alone electronic market, where the vehicle is available to buyers who use the stand-alone electronic market. Variables in the data are described in Table 1.

2.1 Identification of Spatial Arbitrage: Following Overby and Clarke (2012), we identify instances of spatial arbitrage as follows. First, we identify 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 may occur because an arbitrageur is engaging in spatial arbitrage, but they may also occur for other reasons. For example, a buyer may flip a vehicle if he is unable to retail it and chooses to liquidate it in the wholesale market. A dealer may also flip a vehicle after making improvements to it (e.g. repairing dents, replacing tires). 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). 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 7 day interval is reasonable because of the time needed to: a) complete paperwork at the “source” facility where the vehicle was purchased, b) transport the vehicle to the “destination” facility at which it will be sold, and c) register the vehicle for sale at the “destination” facility. A hypothetical example of an instance of spatial arbitrage is as follows: traderID #111 purchased a vehicle with VIN 49567xxxxxxx at the Miami facility on Feb 10, 2003 for \$8,500 and then sold the vehicle at the New Orleans facility on Feb 15, 2003 for \$9,800. The third panel of Figure 1 illustrates trends for the percentage of vehicles that were arbitrated (using $\alpha=7$) over the sample period. Two things are particularly noteworthy. First, spatial arbitrage is rare; there is less than a 1% chance that any given vehicle will be

spatially arbitrated. Second, the prevalence of arbitrage fell over time, although an increasing percentage of arbitrage transactions began with the arbitrageur purchasing the vehicle from the stand-alone electronic market. We explore these trends more formally below.

3.0 Hypotheses

3.1 Webcast channel and spatial arbitrage: Spatial arbitrageurs seek to identify and purchase vehicles that are undervalued at one facility (the “source” facility) and sell them for a profit at a different facility (the “destination” facility). These opportunities exist because buyers at the destination facilities are either unaware of better prices at the source facilities, unable to access these facilities, or both. A key feature of the webcast channel is that it allows buyers to check prices at facilities across the country simply by opening a browser window and observing the auctions via the webcast stream. The webcast channel also allows buyers to purchase vehicles from remotely-located facilities without having to travel. As a result, we posit that implementation of the webcast channel increasingly enabled remotely-located buyers to locate and purchase undervalued vehicles that would otherwise have been purchased by an arbitrageur. As such, we posit that implementation of the webcast channel reduced the prevalence of spatial arbitrage by limiting arbitrageurs’ ability to source vehicles. Because the webcast channel was implemented in phases, we test this by comparing the prevalence of arbitrage between lanes that were webcast-enabled to those that were not (as discussed further below.) This leads to the following hypothesis.

Hypothesis 1 (H1): Vehicles sold in webcast-enabled lanes are less likely to be spatially arbitrated than vehicles sold in non-webcast-enabled lanes.

Based on the same logic, we also test the following hypotheses.

Hypothesis 2 (H2): When sourcing vehicles, arbitrageurs are more likely to choose facilities where the webcast channel is not widely deployed.

Hypothesis 3 (H3): Vehicles sold in webcast-enabled lanes are more likely to be purchased by remotely located buyers than vehicles sold in non-webcast-enabled lanes.

3.2 Stand-alone electronic market and spatial arbitrage: Sourcing undervalued vehicles offered at the physical market facilities may be difficult for arbitrageurs because of the 30-45 second window within which vehicles are auctioned. The short bidding window means that: a) an arbitrageur won't know if a vehicle is undervalued until the bidding begins, and b) he must be actively bidding on the vehicle during the 30-45 second window to purchase it. This is true regardless of whether the arbitrageur is bidding on vehicles offered at physical facilities in person or via the webcast channel. These constraints are relaxed in the stand-alone electronic market, which may make it easier for arbitrageurs to source undervalued vehicles. First, many of the vehicle listings in the stand-alone electronic market include a posted "Buy Now" price. Arbitrageurs can monitor the market for vehicles with unusually low Buy Now prices, purchase them, and profitably arbitrage them. Second, bids may be placed on vehicles in the electronic market over a span of 1-3 days rather a span of 30-45 seconds. As such, arbitrageurs can submit low bids on a high number of vehicles, in the hopes that at least some of those bids will be accepted. Both of these behaviors may yield profitable arbitrage opportunities, and neither are possible for vehicles offered at the physical market facilities. Thus, we posit the following.

Hypothesis 4 (H4): Vehicles sold in the stand-alone electronic market are more likely to be arbitrated than vehicles sold in the physical market.

H1 posits that electronic commerce (in the form of the webcast channel) will decrease the prevalence of spatial arbitrage by helping buyers and sellers trade across geography directly, thereby disintermediating the arbitrageur "middlemen." By contrast, H4 posits that electronic commerce (in the form of the stand-alone electronic market) will *increase* the prevalence of spatial arbitrage by helping arbitrageurs identify mis-priced vehicles. Although these may appear

contradictory, they are reconciled as follows. Support for H1 (with corroborating support from H2 and H3) will reflect an improvement in market integration. Support for H4 will reflect an improvement in arbitrageurs' ability to exploit a given level of market *non*-integration, rather than a decrease in market integration.

4.0 Analysis and Results

4.1 Testing H1: We leveraged the phased implementation of the webcast channel to conduct a natural experiment to test H1 (and H3). Because the webcast technology was deployed at different times for each lane at each facility, we observe many instances in which very similar vehicles were sold at the same facility on the same day, some in lanes that were webcast-enabled and some in lanes that weren't. We considered vehicles sold in non-webcast-enabled lanes to be potential control vehicles and vehicles sold in webcast-enabled lanes to be potential treated vehicles. We matched control vehicles to treated vehicles on facility, sale date, vehicle year, make, model, mileage, valuation, and seller type (see Table 1). If the vehicles are matched adequately, then the only material difference between the control and treated groups are that the control vehicles were sold on non-webcast-enabled lanes while the treated vehicles were sold on webcast-enabled lanes. This approximates a randomized experiment and allows us to attribute any significant differences in transaction outcomes (such as whether a vehicle was later arbitrated) to the "treatment effect" of being sold in a webcast-enabled lane. See (Iacus et al. 2011; Imbens 2004) for a discussion of this type of matching estimation.

We matched vehicles using exact matching and coarsened exact matching ("CEM"). Each treated vehicle could only be matched to a control vehicle with the same *facilityID*, *vehicleYear*, *vehicleMake*, *vehicleModel*, and *sellerType*. In addition, we coarsened *vehicleValuation* and *vehicleMileage* into bins of width 1,000 and 2,000, respectively, and only allowed matches

between vehicles in the same bins. For example, a treated vehicle whose *vehicleValuation* was between \$10,000 and \$11,000 and whose *vehicleMileage* was between 20,000 and 22,000 could only be matched to a control vehicle whose *vehicleValuation* and *vehicleMileage* were within the same ranges (and that was also exactly matched on the variables above)⁵. We also restricted matches to vehicles sold in the same week. Because there were essentially no potential control vehicles available after 2007 (since the webcast channel was almost fully deployed; see Figure 1), we restricted the analysis to observations from January 1, 2003 to December 31, 2007. The matching procedure yielded 45,933 matched strata/cells that each contained at least one treated vehicle and at least one control vehicle. There are 106,950 treated vehicles and 73,633 control vehicles contained within the 45,933 strata in the matched sample. To examine the quality of the matches, we examined the balance between the treated and control observations as follows. First, we calculated the means of *vehicleMileage* and *vehicleValuation* for the treated and control vehicles in each of the 45,933 strata in the matched sample, referred to as the *strata means*. We then used a t-test to test whether these strata means differed significantly between the treated and control groups. As shown in Table 2, the differences are minimal; although the t-statistic for *vehicleValuation* is significant, the average *vehicleValuation* differs by only \$2.17 between the two groups. Overall, we believe that our matches are very precise and satisfy the unconfoundedness condition for valid matching estimation (Iacus et al. 2012; Imbens 2004).⁶

Each vehicle purchased in the market might be subsequently arbitrated. We used the

⁵ Coarsened exact matching temporarily coarsens each chosen variable into bins and exact matches on those bins. CEM then restores the original (not-coarsened) values of the variables for analysis. The CEM matching procedure may match an uneven number of control observations to treated observations. To account for this, CEM generates weights. Using these weights in an estimation procedure (such as a regression model) generates the sample average treatment effect. See Iacus et. al (2011) for details.

⁶ Additional discussion of whether our matching estimation meets the unconfoundedness assumption, including institutional details about how vehicles are assigned to lanes at the physical market facilities, is withheld due to the CIST 15-page limit.

following regression model to test H1: $is_arbitrated_k = \beta_0 + \beta_1 * WebcastEnabled_k + \varepsilon$, where k subscripts the vehicle. $is_arbitrated_k = 1$ if the vehicle was later arbitrated and 0 otherwise. $WebcastEnabled_k = 1$ if the vehicle was sold in a webcast-enabled lane and 0 otherwise. We fitted the model on the matched sample using weighted least squares, with the weights provided by the CEM procedure (see footnote 5). Table 3 shows the results. Our focal results are based on defining spatial arbitrage using the $\alpha=7$ threshold; we varied this threshold for robustness. Given the binary nature of the dependent variable, we also used logistic regression; these results are consistent with the weighted least squares estimates (which are based on a linear specification.)

The intercept (β_0) represents the average probability that a vehicle sold in a non-webcast-enabled lane was later arbitrated. Alternatively, it can be thought of as the percentage of vehicles sold in non-webcast-enabled lanes in the matched sample that were later arbitrated (0.57% with $\alpha=7$). The treatment effect of webcast enablement is captured by β_1 . The effect is negative and significant; vehicles sold in webcast-enabled lanes were approximately 23% less likely to be arbitrated. This indicates that webcast implementation reduced spatial arbitrage, supporting H1.

4.2 Testing H2: For each arbitrage flip, the arbitrageur made a choice to source the vehicle from a particular facility as opposed to some other facility. We examined how they made this choice via a discrete choice model (Train 2003), which allows us to test H2. Discrete choice models are based on the assumption that decision-makers are faced with a set of alternatives and that they choose the alternative that maximizes their utility. Fitting a choice model requires the researcher to define the set of alternatives available to the decision-maker and to specify a utility function for each alternative. We defined the set of alternatives (referred to as the *choice set*) for each arbitrage transaction conducted by arbitrageur i on day t as those facilities at which: a) arbitrageur i ever made a purchase during the sample period, and b) vehicles were auctioned on

day t . We model the utility of each facility j to arbitrageur i at time t as

$$U_{ijt} = \beta_{0,j} + \beta_1 * PctSoldWebcast_{jt} + \beta_2 * Distance_{ij} + \beta_3 * Supply_{jt} + \beta_4 * Supply_{jt}^2 + \beta_5 * MeanPVRatio_{j(t-30)} + \beta_6 * PctSold_{j(t-30)} + \beta_7 * GeoPriceVariability_{j(t-30)} + \beta_8 * NearbyFacilities_j * PctSoldWebcast_{jt} + \varepsilon_{ijt}.$$

We describe the variables in Table 4. We used lagged variables for *MeanPVRatio*, *PctSold*, and *GeoPriceVariability* because the contemporaneous values of these variables are unknown to arbitrageurs when they choose the facility at which to purchase. We use contemporaneous values for the other variables because arbitrageurs either already know them (viz., *Distance_{ij}*) or can calculate them based on the “pre-sale” list of vehicles posted in advance on the intermediary’s web site. We included alternative-specific constants to capture the latent utility of each facility j (represented as the $\beta_{0,j}$ term) and fitted the model using the conditional logit specification. Results appear in Table 5. Our focal results are based on defining spatial arbitrage using the $\alpha=7$ threshold, and we varied this threshold for robustness.

As shown in Table 5, the coefficient for *PctSoldWebcast_{jt}* is negative and significant, indicating that arbitrageurs were less likely to purchase vehicles from facilities at which webcast deployment was high. This supports H2. We used the model estimates to simulate the economic size of this effect as follows. We simulated the percentage change in the number of times an arbitrageur chose facility i when *PctSoldWebcast_{jt}* = 24% vs. when *PctSoldWebcast_{jt}* = 71%, which are the mean values for this variable in 2003 and 2004, respectively.⁷ We did this for each facility. On average, this reduced the probability that an arbitrageur would choose the facility by 30%. This result is consistent with our findings for H1. Webcast enablement reduces the probability that a vehicle was arbitrated (H1), and, by the same token, a high level of webcast enablement makes facilities unattractive source locations for arbitrageurs.

The coefficient for the *NearbyFacilities_j*PctSoldWebcast_{jt}* interaction term is positive and

⁷ This also represents an approximately one standard deviation increase.

significant. This indicates that the disutility of $PctSoldWebcast_{jt}$ is minimized if facility j is surrounded by other market facilities. This is likely because buyers will be less sensitive to $PctSoldWebcast_{jt}$ if a facility is surrounded by several other facilities, because the other facilities represent nearby “destination” locations at which they might complete arbitrage flips.

4.3 Testing H3: As we did for H1, we matched treated vehicles and control vehicles and used weighted least squares to estimate the treatment effect of being sold in a webcast-enabled lane on whether a vehicle was purchased by a remotely-located buyer. As above, t-tests show good balance between the treated and control observations; see Table 6. The regression specification is $remoteBuyer_k = \beta_0 + \beta_1 * WebcastEnabled_k + \varepsilon$, where $remoteBuyer_k = 1$ if the distance between the buyer and the facility at which the vehicle was located was at least one standard deviation above the mean (see Table 1). $remoteBuyer_k = 0$ otherwise. For robustness, we also set $remoteBuyer_k = 1$ if the distance between the buyer and the facility at which the vehicle was located was at least two standard deviations above the mean. Results appear in Table 7.

The results in Table 7 show that webcast enablement increases the probability of a remote buyer by 15%-17%, depending on the measure of a “remote” buyer. This supports H3 as well as our theory underlying the results for H1 and H2. The webcast channel reduces the prevalence of spatial arbitrage by helping remotely-located buyers – who might otherwise be potential “downstream” customers for arbitrageurs – purchase vehicles directly from source locations, thereby disintermediating the arbitrageurs.

4.4 Testing H4: To test H4, we used a similar matching procedure to that which we used to test H1 and H3, 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 or not they were sold in a webcast enabled lane – as potential “control” vehicles. We

ran the matching procedure using the data from 1 Jan 2005 to 31 Dec 2010, as we observe minimal transaction volume in the standalone electronic market prior to 2005. The matching procedure yielded 45,636 matched strata, and the balance between treated and control observations was good, as shown in Table 8.

Similar to H1, the regression specification is: $is_arbitraged_k = \beta_0 + \beta_1 * ElectronicMarket_k + \varepsilon$. We fitted the regression using weighted least squares (with the weights provided by the CEM procedure, as above) and logistic regression. The weighted least squares estimates appear in Table 9; logistic regression results are similar. We used different values of the α threshold for delineating spatial arbitrage for robustness.

β_1 is positive and significant, which supports H4. The results in Table 9 show that vehicles purchased in the electronic market are 73% more likely to be arbitrated than vehicles in the physical market (with $\alpha=7$).

4.5 The overall effect of electronic trading on spatial arbitrage: The results of H1 show that implementation of the webcast channel inhibits spatial arbitrage, and the results of H2 and H3 provide corroborative support. On the other hand, the results of H4 show that the standalone electronic market fosters spatial arbitrage. The negative effect of webcast appears to dominate the positive effect of the standalone electronic market, because the overall percentage of spatial arbitrage decreased over time (see Figure 1). This is likely because there was relatively little transaction volume in the standalone electronic market (see Figure 1). Thus, an increase in arbitrage activity originating in this market cannot compensate for the decrease in arbitrage activity originating in the physical market.

5.0 Conclusion

Spatially-integrated markets are important for the stability and success of the economy, and

e-commerce has the potential to improve market integration by increasing price visibility and facilitating trade across geographic locations. We contribute to research in this area by measuring market integration based on the prevalence of spatial arbitrage, which has several advantages over price-based measures, including accounting for transaction costs and unobserved product quality. Another contribution of our study is that we examine how two different types of electronic commerce influence spatial arbitrage and market integration. We find that the webcast channel decreases spatial arbitrage, reflecting improvements in market integration. We find that the standalone electronic market *increases* spatial arbitrage, reflecting improvements in arbitrageurs' ability to exploit a given level of market *non-integration* (as opposed to a *decrease* in market integration). The overall effect is that e-commerce reduces opportunities for spatial arbitrage and improves market integration.

Our analysis is specific to the wholesale used vehicle market, and the societal implications of market integration in this industry may be less profound than those associated with the integration of markets for staple goods such as corn or wheat. However, well-functioning automotive markets are important in and of themselves, given the surprisingly large impact that the automotive industry has on the overall U.S. economy.⁸ Our results should also generalize to other industries. For example, our conclusion that electronic channels that foster trade across locations improves market integration should generalize to other geographically distributed markets. Similarly, our conclusion that electronic markets that improve arbitrageurs' ability to source undervalued assets foster spatial arbitrage should also generalize.

⁸ In 2011, sales at car dealerships represented approximately 15% of total retail sales in the U.S., and dealership payroll represented approximately 12% of total retail payroll. Source: National Automobile Dealers Association (<http://www.nada.org/Publications/NADADATA/>). Also, almost $\frac{3}{4}$ of oil consumed in the U.S. was consumed by the automotive and transportation industry. Source: U.S. Energy Information Administration.

6.0 References

- Aker, J.C. 2010. "Information from Markets near and Far: Mobile Phones and Agricultural Markets in Niger," *American Economic Journal: Applied Economics*, pp. 46-59.
- Alexander, C., and Wyeth, J. 1994. "Cointegration and Market Integration: An Application to the Indonesian Rice Market," *The Journal of Development Studies* (30:2), pp. 303-334.
- Barrett, C.B. 2005. "Spatial Market Integration," *Entry in Lawrence E Blume and Steven N Durlauf (ed.), The New Palgrave Dictionary of Economics, 2nd Edition (London: Palgrave Macmillan)*.
- Baulch, B. 1997. "Transfer Costs, Spatial Arbitrage, and Testing for Food Market Integration," *American Journal of Agricultural Economics* (79:2), pp. 477-487.
- Iacus, S.M., King, G., and Porro, G. 2011. "Multivariate Matching Methods That Are Monotonic Imbalance Bounding," *Journal of the American Statistical Association* (106:493), pp. 345-361.
- Iacus, S.M., King, G., and Porro, G. 2012. "Causal Inference without Balance Checking: Coarsened Exact Matching," *Political Analysis* (20:1), pp. 1-24.
- Imbens, G.W. 2004. "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics* (86:1), pp. 4-29.
- Jensen, R. 2007. "The Digital Divide: Information (Technology), Market Performance, and Welfare in the South Indian Fisheries Sector," *The Quarterly Journal of Economics* (122:3), pp. 879-924.
- McNew, K. 1996. "Spatial Market Integration: Definition, Theory, and Evidence," *Agricultural and Resource Economics Review* (25), pp. 1-11.
- Minten, B., and Barrett, C.B. 2008. "Agricultural Technology, Productivity, and Poverty in Madagascar," *World Development* (36:5), pp. 797-822.
- Overby, E., and Forman, C. 2011. "The Market Is Flat: How Buyers Use Electronic Channels to Extend Purchasing Reach and Reduce Geographic Price Variance," *Available at SSRN 1400114*.
- Ravallion, M. 1986. "Testing Market Integration," *American Journal of Agricultural Economics* (68:1), pp. 102-109.

7. Figures and Tables

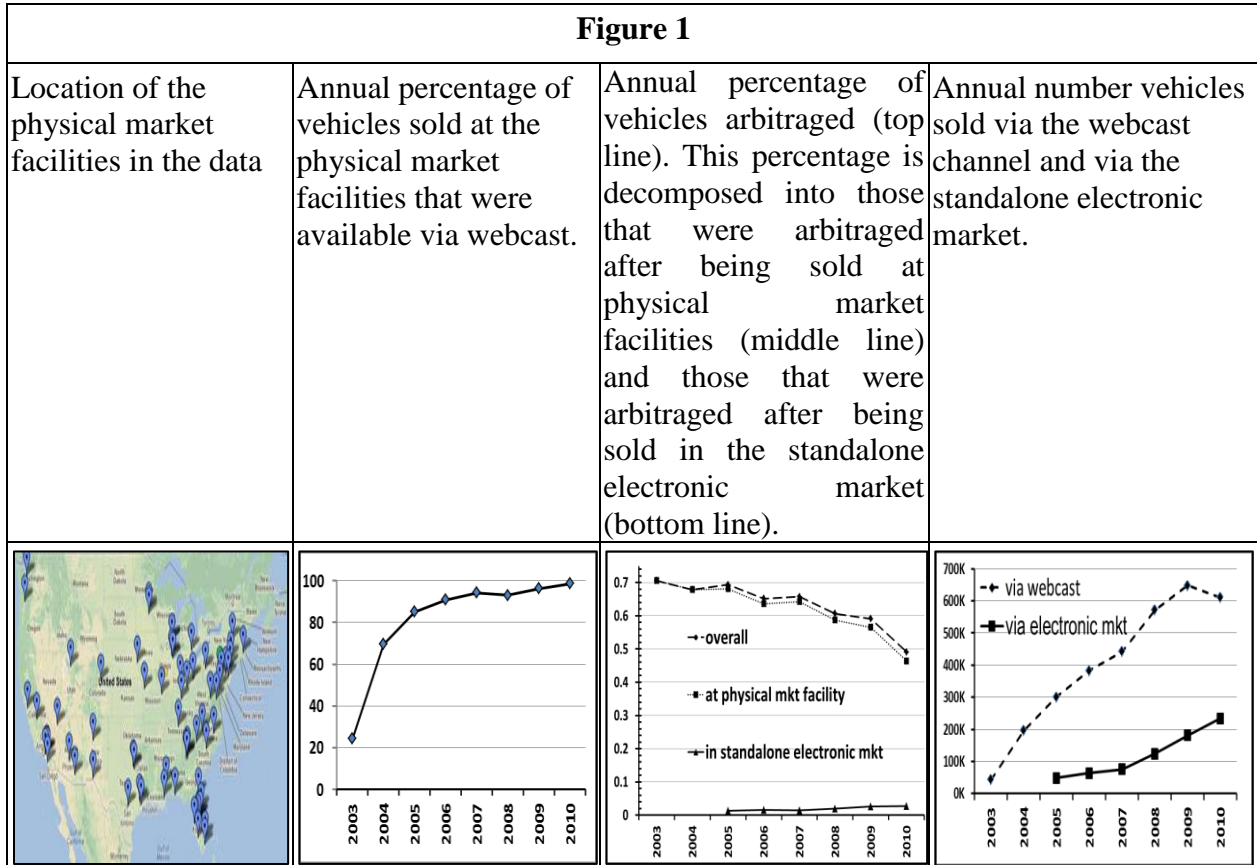


Table 1. Description of variables in the data		
Variable name	Description	Descriptive statistics (mean, Std dev)
<i>facilityId</i>	A unique identifier for the physical market facility.	-
<i>facilityZip</i>	Zip code of the facility.	-
<i>traderId</i>	A unique identifier for the trader. The same ID is used regardless of whether trader is the buyer or seller in a transaction.	-
<i>buyerZip</i>	Zip code of the buyer who purchased the vehicle.	-
<i>saleDate</i>	Date of sale of the vehicle.	-
<i>VIN</i>	The vehicle's unique Vehicle Identification Number (VIN).	-
<i>vehicleMake</i>	The make of the vehicle, e.g. Ford, Honda, etc.	-
<i>vehicleModel</i>	The model of the vehicle, e.g. Taurus, Civic, etc.	-
<i>vehicleYear</i>	The model year of the vehicle	-
<i>vehicleSalesPrice</i>	The vehicle's sales price	Mean- 10435 \$ Std. dev-7913 \$
<i>vehicleValuation</i>	The vehicle's market value as estimated by the intermediary	Mean-11081.8\$ Std. dev- 8527 \$
<i>vehicleMileage</i>	Odometer reading of the vehicle	Mean-64800 Std. dev-49953
<i>electronicMarket</i>	Denotes whether the vehicle was offered in the standalone electronic market or the physical market	1-Standalone electronic market; 0- Physical market
<i>buyerDistance</i>	Distance between <i>buyerZip</i> and <i>facilityZip</i>	Mean 204.6 Std. dev - 343.4
<i>laneId</i>	Denotes the lane at the facility in which the vehicle was auctioned	-
<i>sellerType</i>	Denotes the type of seller (institutional or dealer)	1-institutional; 0-dealer

Table 2. Balance between treated and control observations in the matched sample for testing the treatment effect of the vehicle being sold in a webcast enabled lane on the probability that the vehicle is arbitrated						
Measure	n (strata)	n (treated vehicles)	n (control vehicles)	Mean (treated vehicles)	Mean (control vehicles)	Difference (t-stat)
vehicleValuation	45933	106950	73633	11707.61	11705.44	-2.1736 (-2.374)
vehicleMileage	45933	106950	73633	43984	43955.26	-28.742 (-.9637)

Table 3. Treatment effect of the vehicle being sold in a webcast enabled lane on the probability that the vehicle is arbitrated.				
	Focal result	Robustness results		
	($\alpha=7$ days) is_arbitrated _k	($\alpha=5$ days) is_arbitrated _k	($\alpha=15$ days) is_arbitrated _k	($\alpha=28$ days) is_arbitrated _k
webcastEnabled _k (β_1)	-0.0013 ^{***} (0.0003)	-0.0008 ^{**} (0.0003)	-0.0033 ^{***} (0.0005)	-0.0041 ^{***} (0.0006)
Intercept (β_0)	0.0057 ^{***} (0.0003)	0.0034 ^{***} (0.0002)	0.0118 ^{***} (0.0004)	0.0165 ^{***} (0.0004)
<i>n</i>	180583	180583	180583	180583

Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)
 α is the number of days used to delineate spatial arbitrage.

Table 4. Variables used in the choice model of where arbitrageurs source vehicles. Includes both chosen and non-chosen alternatives.		
Variable name	Description	Mean (Std. Dev)
PctSoldWebcast _{jt}	The percentage of vehicles sold at facility <i>j</i> on day <i>t</i> that were sold in webcast-enabled lanes.	0.62 (0.411)
Distance _{ij}	Distance between facility <i>j</i> 's zip code and arbitrageur <i>i</i> 's zip code.	535.23 (528.18)
Supply _{jt}	The number of vehicles offered at facility <i>j</i> on day <i>t</i> .	783.40 (765.07)
MeanPVRatio _{j(t-30)}	The mean price to valuation ratio of the vehicles sold at facility <i>j</i> in the 30 days prior to day <i>t</i> .	0.98 (0.22)
PctSold _{j(t-30)}	The percentage of vehicles sold at facility <i>j</i> in the 30 days prior to day <i>t</i> .	0.605 (0.12)
GeoPriceVariability _{j(t-30)}	The average geographic price variability (measured as the standard deviation) of the vehicles offered at facility <i>j</i> in the 30 days prior to day <i>t</i> .	1536.77 (242.00)
NearbyFacilities _j	The number of facilities within 350 miles of facility <i>j</i> .	20 (8)

Table 5. Results of the choice model of where arbitrageurs source vehicles				
	Focal result	Robustness check		
	($\alpha=7$ days)	($\alpha=5$ days)	($\alpha=15$ days)	($\alpha=28$ days)
PctSoldWebcast _{jt} (β_1)	-0.9235*** (0.0413)	-0.7242*** (0.0955)	-0.7685*** (0.0387)	-0.7258*** (0.0381)
Distance _{ij} (β_2)	-1.1745*** (0.0136)	-2.4468*** (0.0419)	-1.0715*** (0.0126)	-1.0531*** (0.0124)
Supply _{jt} (β_3)	3.5369*** (0.0213)	3.4765*** (0.0451)	3.1992*** (0.0192)	3.1185*** (0.0188)
Supply ² _{jt} (β_4)	-0.5488*** (0.0053)	-0.5546*** (0.0109)	-0.5001*** (0.0049)	-0.4880*** (0.0048)
MeanPVratio _{j(t-30)} (β_5)	-0.0002 (0.0032)	0.0067 (0.0047)	-0.0009 (0.0031)	-0.0014 (0.0032)
PctSold _{j(t-30)} (β_6)	-2.2461*** (0.0588)	-2.3587*** (0.1280)	-1.8935*** (0.0541)	-1.8273*** (0.0532)
GeoPriceVariability _{j(t-30)} (β_7)	-0.2405* (0.0984)	-0.4594* (0.2188)	-0.2948*** (0.0864)	-0.2558** (0.0836)
NearbyFacilities _j * PctSoldWebcast _{jt} (β_8)	0.0286*** (0.0017)	0.0277*** (0.0040)	0.0257*** (0.0016)	0.0246*** (0.0016)
Facility($\beta_{0,j}$)	Included	Included	Included	Included
Log Likelihood	-122924.36	-24729.127	-137788.44	-140958.85
Number of choices ^a	122358	29982	129128	130385
Alternatives per choice set (Min, Mean, Max)	2,6.1,48	2,5.6,48	2, 6.2,49	2,6.2,49

Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)
 α is the number of days used to delineate spatial arbitrage.
To ensure that coefficients are of similar magnitude for reporting purposes we divided *Supply*, *Distance* and *GeoPriceVariability* by 1000.
^a The number of choices for $\alpha=7$, 15, and 28 is similar. This is because the number of choices reflects distinct instances in which an arbitrageur sourced vehicles from facility j on day t . If he flips at least one of the sourced vehicles within 7 days, then that instance will appear in the $\alpha=7$ estimation. The only time an instance will appear in the $\alpha=15$ estimation but not the $\alpha=7$ estimation is if the arbitrageur flips at least one sourced vehicle within 15 days but none within 7 days, which is relatively rare.

Table 6. Balance between treated and control observations in the matched sample for testing the effect of webcast enablement on the probability that the vehicle is bought by a buyer who is at a distance greater than one and two standard deviations above the mean

Measure	n (strata)	n (treated vehicles)	n (control vehicles)	Mean (treated vehicles)	Mean (control vehicles)	Difference (t-stat)
vehicleValuation	35317	66293	47140	12200.79	12203.65	-2.86 (-2.583)
vehicleMileage	35317	66293	47140	37974.17	37992.44	-18.26 (-0.574)

Table 7. Treatment effect of webcast enablement on the probability that the distance between buyer and facility is greater than 1 and 2 standard deviations above the mean

	remoteBuyer _k (distance > mean + std. dev)	remoteBuyer _k (distance > mean+ 2*std. dev)
WebcastEnabled _k (β_1)	0.0180 ^{***} (0.00219)	0.0151 ^{***} (0.0018)
Intercept(β_0)	0.146 ^{***} (0.00168)	0.0914 ^{***} (0.00139)
<i>N</i>	113433	113433
Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$)		

Table 8. Balance between treated and control groups in the matched sample for testing the treatment effect of selling a vehicle in the standalone electronic channel on the probability that it is arbitrated

Measure	n (strata)	n (treated vehicles)	n (control vehicles)	Mean (treated vehicles)	Mean (control vehicles)	Difference (t-stat)
vehicleValuation	45636	57722	115855	14750.71	14749.72	-0.9887 (-1.27)
vehicleMileage	45636	57722	115855	25517.04	25522.56	5.5197 (1.630)

Table 9. Treatment effect of selling a vehicle in the standalone electronic channel on the probability that it is arbitrated

	Focal result	Robustness result		
	($\alpha=7$ days) is_arbitrage d_k	($\alpha=5$ days) is_arbitrated k	($\alpha=15$ days) is_arbitrated k	($\alpha=28$ days) is_arbitrated k
ElectronicMarket _k (β_1)	0.0028 ^{***} (0.0004)	0.0028 ^{***} (0.0003)	0.0041 ^{***} (0.0005)	0.0049 ^{***} (0.0006)
Intercept(β_0)	0.0047 ^{***} (0.0002)	0.0025 ^{***} (0.0002)	0.0099 ^{***} (0.0003)	0.0146 ^{***} (0.0004)
N	173577	173577	173577	173577
Standard errors in parentheses (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$) α is the number of days used to delineate spatial arbitrage.				