RAND Journal of Economics
Vol. 51, No. 3, Fall 2020
pp. 905–933



# Intermediaries and product quality in used car markets

**Gary Biglaiser***

**Fei Li***

**Charles Murry****

**and**

**Yiyi Zhou*****

*We present empirical evidence supporting that used cars sold by dealers have higher quality: (i) dealer transaction prices are higher than unmediated market prices, and this dealer premium increases in the age of the car as a ratio and is hump-shaped in dollar value, and (ii) used cars purchased from dealers are less likely to be resold. In a model, we show that these empirical facts can be rationalized either when dealers alleviate information asymmetry, or when dealers facilitate assortative matching. The model predictions allow us to distinguish these two theories in the data, and we find evidence for both.*

## 1. Introduction

■ Most transactions are made through a variety of intermediaries such as retailers, dealers, and brokers. Since there is no place for intermediaries in Arrow–Debreu's highly stylized world, to understand the ubiquitousness of intermediaries, one must count on market frictions. One obvious rationale is offered by Rubinstein and Wolinsky (1987): intermediaries can facilitate searching and matching between parties in decentralized markets. Moreover, when goods are heterogeneous, and tastes of agents are idiosyncratic, intermediaries could also improve the allocation or match efficiency. Another popular justification of intermediaries relies on frictions due to an informational asymmetry between agents. As Biglaiser (1993) and Lizzeri (1999) argue, intermediaries can serve as information intermediaries, or certifiers, in markets where there are

* University of North Carolina at Chapel Hill; gbiglais@email.unc.edu,lifei@email.unc.edu.
** Boston College; charles.murry@bc.edu.
*** Stony Brook University; yiyi.zhou@stonybrook.edu.
This article supersedes the previous work titled "Middlemen as Information Intermediaries: Evidence from Used Car Markets." The authors thank Eric Bond, Liran Einav, Igal Hendel, Brad Larsen, Qihong Liu, Alessandro Lizzeri, Brian McManus, Peter Newberry, John Rust, Tobias Salz, Henry S. Schneider, Karl Schurter, Andrei Shleifer, Shouyong Shi, Senay Sokullu, Randy Wright, Andy Yates, Jidong Zhou, editor Katja Seim, anonymous referees, and participants of many conferences and seminars for helpful comments.

motives for adverse selection or consumer sorting. The idea is that intermediaries have more advanced technology and experience to distinguish product quality, so goods traded through them are of higher quality than those traded directly between sellers and buyers. Although the theoretical literature has proposed many distinct rationales for intermediaries, empirical research is limited and focuses almost exclusively on how intermediaries alleviate search frictions. The goal of this article is to examine the role of used-car dealers more comprehensively. We provide evidence supporting the hypothesis that car dealers provide high-quality cars for consumers, motivated either by information asymmetries or by match efficiency motives.

Using administrative registration records of used car transactions from two large states, we examine the prices and resale patterns of cars sold through dealers and cars sold directly by sellers. First, we document a dealer price premium: for the same type of car, transaction prices of dealer sales are higher than transaction prices in the unmediated market.[1] Second, we show that the dealer premium, in dollars, is hump-shaped in car age and as a ratio is increasing in the car age. Third, we document that used cars purchased from dealers are less likely than unmediated transactions to be resold within a short time after the initial transactions. Although there may be many sources of the dealer premium, we argue that these observations are consistent with the hypothesis that part of the dealer premium is due to dealers offering superior cars.

We formalize our dealer quality premium argument with a parsimonious theoretical model to understand a dealer's role in a market with a depreciating good that may experience a failure, or in Akerlof's parlance, become a lemon. When selling a car, a seller can visit a dealer, who decides how much, if anything, to offer for the car. The seller can either trade with the dealer or go to the unmediated market and sell the car directly to buyers.

Based on these ingredients, we show that two prevailing theories about intermediaries can explain empirical observations. First, we assume that the seller privately observes the quality of the car, but dealers are experts who can run a test to ascertain quality. The unmediated market understands that the dealer is an expert and has reputation concerns; therefore, dealers trade higher-quality cars on average and enjoy a price premium over the unmediated market. Since the age, or vintage, of a car represents the hazard of the car becoming a lemon, it has an important effect on the dealer's price premium. On the one hand, the dealer's information role increases as the car ages, but on the other hand, even a high-quality car depreciates as it ages. We show that this leads to a dealer premium pattern described above. Besides, the dealer's expertise in screening car quality generates a selection mechanism: cars purchased through dealers are more likely to be of high quality than those purchased from sellers directly. By the classic adverse selection logic, lemons will be resold sooner than high-quality cars.

Second, we consider a model with complete information but consumer heterogeneity: buyers have either a high or low valuation. In this case, dealers serve as a platform to facilitate assortative matching between buyers and sellers in the presence of search frictions. In equilibrium, dealers only sell high-quality cars and attract high-valuation buyers. In the unmediated market, cars with both high and low quality are sold, and low-valuation buyers purchase these cars. The dealers' price premium is justified by the matching efficiency they create. We show that the age profile of a dealer's price premium is also consistent with the data. Also, the initial allocation is inefficient in the unmediated market: buyers with low valuations may purchase high-quality cars, giving them an incentive to resell their cars. Therefore, the model also predicts that the resale rate is higher when the car is purchased from the unmediated market than when it is purchased from a dealer.

After presenting the theory, we turn back to the used car data to distinguish between the asymmetric information and sorting theories. Distinguishing these two theories is essential for two reasons. First, it allows us to identify the source of the inefficiency in the used-car market, by better understanding the role of dealers in this market. For example, imagine that a dealer sells a

[1] Throughout the article, we will refer to cars sold by dealers as mediated transactions and cars sold by sellers to buyers without going through a dealer as direct or unmediated transactions.

car with a warranty contract. If consumers perfectly observe the car's quality, then the warranty is simply an insurance contract; if consumers do not observe the car's quality, the warranty can be viewed as a collateral contract that enables the dealer to commit to selling high-quality cars. Second, as we will discuss further in the literature review section, it has been debated whether information asymmetry exists between buyers and sellers of used cars and whether there should be some regulation in this market; for example, whether there should be a "lemon's laws." Therefore, identifying the source of market inefficiency also has important policy implications.

Specifically, the two theories make different predictions about the type of resold cars. In the asymmetric information theory, the expected quality of resold cars is lower than the expected quality of cars that are initially transacted because the resales are driven partly by buyers who want to get rid of lemons. On the other hand, in the sorting theory, the expected quality of resold cars is higher because the resales take place to improve the initial allocations. We show that the resale rate is increasing in car age until age 9 at a faster rate for cars purchased from sellers directly than cars purchased from dealers, and the resale price is more likely to be lower than the initial transaction price. Both of these facts support the presence of asymmetric information. However, the tests do not rule out the sorting story, and we find evidence that the sorting mechanism tends to be more important for older vintages. Intuitively, when a car is relatively old, its quality is more likely to be public information but more heterogeneous, so car dealers' role in sorting is more critical as a car ages.

Many factors make the used car market suitable for our study. First, cars are complicated machines that require specialized care and maintenance; dating back to Akerlof (1970), the used car market has been showcased as an example of a market rife with information asymmetries—sellers have more information about the product's quality than buyers do. Second, the unmediated market is highly decentralized, and used cars are heterogeneous, making the search and matching frictions in the unmediated market non-trivial. Third, there are thousands of dealers per state, but still many transactions are not intermediated, allowing us to examine the difference of their transaction patterns. Fourth, the used car market is large, with the total retail sales over 500 billion dollars annually in the United States.[2] In 2016, 38.5 million used cars were sold in the United States, more than twice the number of new cars sold.[3] Last, dealers are very active participants in the used car market. Nationally, about two-thirds of used car sales are made by dealers, and the other one-third occur directly between sellers and buyers. There are important differences between unmediated sales and dealer sales. Unmediated sales are much less regulated than dealer sales. Dealers are long-run players who sell many cars and care about their reputations, whereas sellers in the unmediated market participate very infrequently and do not have reputation concerns. Furthermore, dealers are experts who may transact the same type of cars many times and who employ mechanics on site.

We present empirical evidence of the quality provision role of intermediaries consistent with a model that reflects features of the used car industry. There has been growing interest from empirical researchers in analyzing the role of intermediaries, but most of these studies focus on intermediaries' roles of resolving search frictions. Recent examples include Hendel, Nevo, and Ortalo-Magné (2009), Gavazza (2016), Salz (2017), and Donna et al. (2018). One exception is Galenianos and Gavazza (2017), who estimate a model of cocaine buyers and sellers and show that reputation concerns help support an equilibrium where the dealer offers high-quality drugs in the presence of asymmetric information. However, unlike in their setting where trades only happen through dealers, both dealers and individuals facilitate trades in the used car market,

[2] This number, constructed from Edmunds' and Manheim's yearly reports, represents revenues from franchised and independent dealers only, so it is a conservative reflection of the size of the industry. We found conflicting reports about the total revenues of the direct party sector.

[3] Our general understanding of the industry is conversations with dealers and from various industry reports, including Edmunds' "Used Vehicle Market Report," Manheim's "Used Car Market Report," and Murry and Schneider (2015). For industry reports, see dealers.edmunds.com/static/assets/articles/2017_Feb_Used_Market_Report.pdf and publish.manheim.com/content/dam/consulting/2017-Manheim-Used-Car-Market-Report.pdf

and we examine both the information asymmetry story and the assortative matching story. Our work is closely related to many studies on adverse selection, sorting and market segmentation, and intermediaries.

The theoretical foundations of this article lie in the work of three strands of literature about intermediaries. First, Biglaiser (1993), Biglaiser and Friedman (1994), and Biglaiser and Li (2018) argue that in an environment with asymmetric information *a la* Akerlof (1970), intermediaries emerge to identify lemons. Second, there is a large literature discussing the function of intermediaries to save search costs of agents in the unmediated market; see Rubinstein and Wolinsky (1987), Gehrig (1993), Yavaş (1994, 1996), Spulber (1996), Rust and Hall (2003), Wright and Wong (2014), Nosal, Wong, and Wright (2015, 2017), Rhodes, Watanabe, and Zhou (2018) as examples. Last, there is a literature emphasizing the role of intermediaries in facilitating allocation efficiency. Biglaiser and Friedman (1999) point out that in the presence of asymmetric information, intermediaries can facilitate market segmentation and improve social welfare. Johri and Leach (2002) and Shevchenko (2004) consider economies with search frictions, a variety of goods, and agents with heterogeneous tastes. By holding a large number of inventories, an intermediary can increase the probability to satisfy the demand of random customers. Although an intermediary's aforementioned roles have been well recognized on the theoretical side, the literature on the empirical side almost exclusively emphasizes that intermediaries save search costs. One exception in addition to Galenianos and Gavazza (2017) discussed above is Peterson and Schneider (2014), who report that cars sold by dealers require fewer repairs than cars sold by sellers, although this is not their primary focus. Gavazza (2016) shows that dealers reduce trading frictions through costly intermediation, but also impose an externality by crowding out the number of unmediated transactions. In other industries, Hendel, Nevo, and Ortalo-Magné (2009) compare house sales on a For-Sale-By-Owner online platform to the Multiple Listing Service, and Salz (2017) investigates intermediaries' role in relieving search costs in New York City's waste disposal market. Our contribution to this literature is twofold. First, we empirically test whether an intermediary provides high-quality products. Second, we propose tests to empirically distinguish the aforementioned competing theories about the roles of intermediaries.

Inspired by Akerlof (1970), economists have long studied whether information asymmetry exists in the leading example of a lemon market, the used car market. However, by definition, asymmetric information can hardly be directly measured, so economists turn to test its implication: adverse selection. The evidence about adverse selection is mixed: some find evidence of adverse selection; others do not. See Bond (1982, 1984), Lacko (1986), Genesove (1993), Engers, Hartman, and Stern (2009), and Adams, Hosken, and Newberry (2011) as examples. We contribute to the literature by comparing the transaction price, conditional on car age, of dealers with those in unmediated sales. Also, rather than testing for the presence of asymmetric information by examining sellers' adverse selection, we focus on the selection made through dealers. There is also a large literature on asymmetric information in insurance markets, including Cardon and Hendel (2001), Einav, Finkelstein, and Schrimpf (2010), and Hendren (2013). It is likely that the role of intermediaries that screen asymmetric information is very limited in these markets, as buyers tend to screen sellers on a rich set of observable characteristics, like demographic information and credit scores.

Recently, inspired by the test derived by Hendel and Lizzeri (1999), Peterson and Schneider (2014) consider a car as an assemblage of parts, some with asymmetric information, and others without, and find evidence of adverse selection and consumer sorting. Our article not only confirms Peterson and Schneider (2014)'s finding that both adverse selection and consumer sorting exist and also identifies the group of cars in which the information asymmetry story plays a more important role.

The idea of using turnover rates to proxy quality has been widely used in the literature. For example, in the period prior to the 2007 financial crisis, securitized mortgages had significantly higher default rates than loans originated and held by the same institution, which is regarded as evidence of adverse selection. See Berndt and Gupta (2009), Mian and Sufi (2009), and Keys et al. (2010). We focus on resales taking place within a short time period after the initial transactions to tease out the reallocation resulting from depreciation that plays a central role in Bond (1983), Hendel and Lizzeri (1999), and Peterson and Schneider (2014).

The rest of the article is organized as follows. Section 2 documents that the pattern of price premium and resale rates in the data are consistent with the hypothesis that car dealers sell cars with higher quality than the unmediated market. In Section 3, we develop a theoretical model and show that the empirical regularities can be explained either when dealers alleviate asymmetric information or when they facilitate assortative matching. Section 4 empirically distinguishes the two theories. Section 5 concludes. We relegate additional empirical results and discussions of the model to online appendix.

## 2. Evidence of Quality Difference

■ In this section, we investigate whether there is empirical evidence consistent with car dealers selling higher-quality products than those sold directly on the unmediated market. We center our empirical analysis on two relationships in our data. First, we focus on the relationship between the dealer price premium and car age because the quality of a car changes with age. Second, we turn our attention to the frequency at which cars are resold immediately.

Before we explain our empirical strategy, it is important to define the word "quality." As cars age, various features will age differentially from car to car due to both underlying differences in parts of the car that are unobserved at production and the differences in how the cars are driven and maintained by owners. Some users add value to cars (or substantially slow down depreciation) by performing extra maintenance or adding features like paint coating or improved interior features. On the other hand, some users do not perform regular maintenance or may wear the interior or exterior of the car due to their driving habits. These features, which are typically unseen by the researcher but are valued by the consumers, are what we consider to be a car's *quality*. Importantly, fixing a car, we do not consider observable features like car age (directly), mileage, or make/model/trim as part of the car's quality per se, but these features may determine how quality evolves in the future.

It is important to note that our definition of quality, although not observed by us, may be public information between sellers and buyers or may be private information of sellers. For example, a car may have visible exterior/interior damage or the owner may have receipts from maintenance, oil changes, or professional detailing. Alternatively, there may be wear in the engine or drivetrain that would be difficult for a non-expert to detect, or the current owner may hide maintenance records that contain information about recurring problems due to a defect.[4] Our first goal is to present evidence that used car dealers offer higher-quality cars than sellers in the unmediated market.

□ **Empirical strategy.** The challenge confronted by all studies is that, by definition, quality is unobservable to the researcher and therefore hard to measure directly. In the literature, quality measures are typically indirect data suggested by the insights from economic theory. Our approach is to examine two features of used car markets that play a prominent role in the existing literature: prices and resale rates. We formalize the following theoretical arguments in Section 3.

First, we examine the *age pattern* of the price premium of dealer sales over unmediated sales. Hayek (1945) suggests that the market price aggregates dispersed information and reflects the expected value of traded products. This insight has been formalized in various settings. Motivated by this literature, the starting point of our empirical strategy is based on the specification

[4] Peterson and Schneider (2014) elaborate on this distinction between observed and unobserved quality using repair services for particular parts of the car.

that the transaction price is positively correlated with the car quality.[5] In the used car market, if dealers sell higher-quality cars than the unmediated market, one should expect that dealers enjoy a positive price premium relative to the unmediated market for cars with observably identical characteristics. However, a positive price premium does not necessarily imply that dealer cars have superior quality because, besides selling the product, dealers provide a sequence of pretransaction services such as search cost savings, financing, explicit warranties, and positive shopping experience (from knowledgeable product discovery). These services have nothing to do with the quality of the product per se, but they do affect the buyers' payoff and therefore their shopping decision and willingness to pay.[6]

To isolate the effect of the quality premium, we examine the effect of car age on the price premium. If the price premium can be partially attributed to the quality premium, it should vary across the vintage of cars. The logic is simple. (i) The value of a car depreciates over time regardless of its quality, which suggests that the difference between high-quality cars and low-quality cars, and therefore the price premium, should fall as a car ages. (ii) It is natural to believe that an older car is more likely to be of low quality; or in other words older cars are more likely to suffer a defect or have visible wear, resulting lower *expected quality* of older cars. This effect suggests that the dealer's value-added by providing high-quality cars, rewarded by price premium, should increase in car age. However, it is difficult to use the value-added of the dealer's pretransaction service to generate the age pattern of the price premium.

Second, we examine the *post-transaction* resale rate of cars. If dealers sell higher-quality cars, cars should be quickly resold less often than if they are from the unmediated market. The reason is twofold. (i) When a buyer purchases a car she may be uncertain about the quality of the car and when she realizes that her purchase is a lemon, she will be more likely to resell it. (ii) The initial allocation in the unmediated market may be less efficient than through dealers due to buyer heterogeneity in willingness to pay for quality and differential search and matching frictions. On the one hand, dealers chiefly trade high-quality cars at higher prices, which mainly attract buyers with high valuations, leading to more efficient allocations. On the other hand, in the unmediated market, transactions are less organized, information is less aggregated, and car quality is more dispersed, so inefficient allocations are more likely to occur. In this case, reallocations take place to "correct" the initial allocations from the unmediated market. A buyer who purchased from the unmediated market will be more likely to meet another buyer who has a higher valuation for the car, leading to higher resale rates for cars traded in the unmediated market.

□ **Price premium of dealers.** We analyze the universe of used car transactions registered in Virginia from 2007 to 2014 and document the difference in transaction prices between dealer sales and unmediated sales, and the car age patterns of this difference. The dataset is obtained from the Virginia Department of Motor Vehicles (VA data). For each registration, we know the transaction date, price,[7] the first 12 digits of the Vehicle Information Number (VIN) which is a unique number assigned to a vehicle that contains information to describe and identify the vehicle,[8] and odometer mileage. We also know some information about the buyers and sellers. Sellers are either marked as "private sellers," or as dealers with a dealer identification number.

[5] Notice that the dispersed information may not be perfectly aggregated by the market price in the presence of search frictions, asymmetric information and common value assumptions. Our specification relies on the assumption that the market price (at least partially) aggregates dispersed information.

[6] Many other factors may also contribute to the price premium, for example, (i) underreported price in unmediated transaction for tax avoidance, (ii) bargaining power difference between dealers and sellers. What is important for our empirical strategy is that these other factors do not correlate with the age of a car in the particular way our model predicts.

[7] The price of a car is the transaction price reported to the state for tax purposes. Car dealers sometimes offer a car as "certified pre-owned"(CPO). In these cases, the price also includes any benefits from CPO status. According to *Edmunds*, about 7% of all used car transactions are CPO cars.

[8] The VIN standard, created by the National Highway Traffic Safety Administration (NHTSA) and enforced starting with the model year 1981, is required of all vehicles manufactured for use in the United States.

**TABLE 1 Descriptive Statistics of VA Data**

| | Unmediated Sales | | | Dealer Sales | | |
|---|---|---|---|---|---|---|
| | Mean | SD | Median | Mean | SD | Median |
| Price | $3,943 | $5,043 | $2,000 | $12,849 | $8,316 | $11,995 |
| Mileage | 131,871 | 57,279 | 131,960 | 77,470 | 49,901 | 67,529 |
| Car Age | 11.09 | 4.35 | 11 | 6.01 | 4.02 | 5 |
| **Total Transactions:** 5,301,157 | | | **Dealer Sales:** 60.13% | | | |

Note: The VA data include all used car transactions registered in Virginia from January 1, 2007, to December 31, 2014. Sample selection is described in text. Data source: Virginia Department of Motor Vehicles.

We merge the dealer identification numbers with a separate dataset provided by the VA data that includes identification numbers matched to dealer names and addresses. The zip codes of buyers are also provided for many, but not all, observations. The zip codes of private sellers are also provided, but for many fewer transactions than for buyers.

Based on the information provided by the edmunds.com API, we decode the "squish VINs," the first 12 digits of the VINs except for the ninth digit, into the make, model year, model, and exact trim with a particular set of options. The trim is a specific configuration of engine and other options available for a car. Most popular models have at least two trims available. For example, the squish VIN of 4T1BF3EKBU identifies a 2011 Toyota Camry LE with a 4-cylinder engine and an automatic 6-speed transmission. Using the zip codes of buyers and sellers, we merge the VA data with a list that matches zip codes to counties and demographic information.

We make a number of sample selection decisions for the raw data to focus on our research questions and to clean obvious mistakes in the raw data. First, we drop 387,926 transactions when dealers are buyers because our focus is the retail market. Second, we discard 209,124 transactions with negative odometer readings or car ages more than 20 years old. We also discard transactions with recorded prices less than $500 or greater than $50,000. These transactions are outliers (for example transactions between family members) or were mistakenly recorded. Last, we exclude those extremely unpopular products with fewer than 100 transactions over the 8 years (from 2007 to 2014) which account for less than 2% of the sample. In the end, we are left with 5,301,157 transactions, representing 35,248 unique model–model year–trims. Among them, 3,187,470 transactions (60%) were sold by dealers and the remaining 2,113,687 (40%) were sold directly by owners.

Table 1 presents summary statistics from our sample, including the transaction price, car age, and odometer mileage separately for unmediated sales and for dealer sales. Overall, cars sold by dealers were substantially newer and more expensive than those directly sold by sellers. Specifically, an average dealer car was around 6 years old and sold at a price of $12,849, whereas an average non-dealer car was 11 years old and sold at a price of $3,943. However, the standard deviations of car age and transaction price are large, indicating that there was substantial heterogeneity across transactions for both segments.

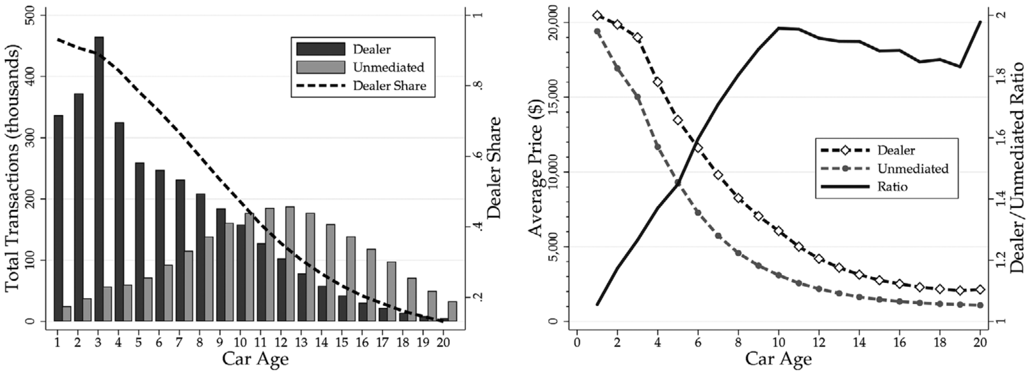
Figure 1a presents the total transactions of the two segments across different car vintages. First, the total number of dealer sales falls in car age after peaking at 3-year-old cars, which is the common lease length for leasing cars. Second, the total number of unmediated sales increases in car age until age 12 and then falls in car age. We also graph the share of dealer sales by vintage, which is strictly decreasing with car age.[9] In addition, we merge our transaction data with Census data to get the local demographics at the buyer's zip code.

Last, we summarize the prices of dealer sales and unmediated sales for every car age. We plot the average transaction price by car age in Figure 1b. The two downward-sloping lines are the transaction prices for dealer sales and unmediated sales. The upward-sloping line (associated

[9] These patterns continue to hold, on average, after controlling for car make and model effects, implying that these patterns are not the product of compositional effects in the type of cars sold across seller types and vintage.

FIGURE 1

DEALER SALES AND UNMEDIATED SALES IN THE VA DATA




(a) Total Transactions (b) Mean Transaction Prices

Note: An observation is a single used-car transaction in Virginia from 2007 to 2014. The sample is described in the text. Data source: Virginia Department of Motor Vehicles.

with the right axis) is the ratio of these two prices. The average prices of dealer sales are higher than unmediated sales at every age. The *price difference* increases at first, and then decreases, implying that very old cars have similar average prices. The *price ratio* is increasing until age 10, and then flattens out. These age patterns are the primary motivation for the remainder of our empirical analysis on the dealer premium. Of course, prices of dealer sales and unmediated sales may differ across vintages due to compositional effects, and the following empirical analysis will control for these compositional changes by using within trim variation in prices. In the remainder of the empirical analysis, we examine how prices are correlated with car age, but it could also be the case that mileage is the primary consideration when thinking about the quality of a car. Age and mileage are highly correlated, with a correlation coefficient of 0.70 in our sample. Both variables also have broadly similar patterns with respect to transaction prices. The dealer premium has a similar relationship to mileage (not shown in the graph).

We define the dealer price premium formally as it relates to our data. The price premium is the average difference between the dealer price and the price in the unmediated market, conditional on observed car characteristics (observed by the researcher) including the "type" of car and mileage. We define a "type" of car as a unique make, model, model year, and trim. We also consider the price premium ratio, which is the average ratio of dealer prices to direct prices, conditional on observable car characteristics. To estimate the dealer premium, we estimate a hedonic price regression where we regress log price on various transaction characteristics including car mileage, month, and year effects, an indicator for dealer, indicators for different car ages, and age indicators interacted with the indicator of dealer. Importantly, we difference out any observed characteristics of cars by including *type* (make-model–model year–trim) fixed effects. The coefficients before the interaction terms of the dealer and car age indicators capture to what extent the dealer price premium co-varies with car age. Our strategy is to compare prices of two observationally equivalent cars (same model, same model year, same trim, same odometer mileage, and vintage), with one being sold by a dealer and the other one being sold directly by the owner, and we examine how this price difference varies in car age.

In specification (1), we use the full sample. To relieve the concern that new car dealers may take into account the substitution between their new cars and used cars when they price their used cars (as well as issues with CPO designated cars discussed above), in specification (2) we limit our analysis to unmediated sales and dealer sales from used-car-only dealers who

**TABLE 2 Results of Dealer Premium Regressions Using the VA Data**

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| log(Mileage) | −0.286 | −0.326 | −0.311 | −0.375 |
| Constant | 12.553 | 12.904 | 12.736 | 13.098 |
| Age Indicators | | ………….. See Figure 2a ………….. | | |
| Age–Dealer Interactions | | ………….. See Figure 2b ………….. | | |
| $R^2$ | 0.750 | 0.471 | 0.547 | 0.460 |
| Num. Observations | 5,301,157 | 3,578,513 | 1,151,447 | 4,067,915 |

Note: An observation is a single transaction from the sample described in the text. The dependent variable is the log of transaction price, and all specifications include product (make–model–model year–trim) fixed effects, log of the odometer mileage, month and year dummies, car age indicators, and interactions of age indicators and dealer indicator. All point estimates are statistically significant at least at 1% significance level. Specification (1) includes the full sample. Specification (2) excludes cars sold by new car dealers. Specification (3) includes popular car models only. Specification (4) excludes cars younger than 4 years old.

do not have new car franchises. Unpopular products may also have liquidity issues which may affect their prices and induce correlation between search rents and car age. For example, older desirable cars may have excess demand. To relieve this concern, in specification (3) we include only the most popular car types that have more than 10,000 sales during the sample period. Last, to reduce the potential impacts of leasing cars, rental cars, CPOs, and substitution from new cars, in specification (4), we include only transactions of cars at least 4 years old. The statistics of the samples used in the four specifications, including the price, car age, and odometer mileage for the two segments are relegated to online appendix.

The estimation results are reported in Table 2 and Figure 2. The estimates are extremely precise, with every coefficient we report being statistically significant at least at the 1% significance level, using robust standard errors. As expected, the coefficient for the log of mileage is negative. The coefficients and associated standard errors for car age indicators are reported graphically in Figure 2a. The car age coefficients are all negative and monotonically decreasing with age, implying that older cars are valued less. Notice that the age coefficients for specification (4) are above those for other three specifications. This is because in specification (4) the baseline age is 4 years old rather than 1 year old in other specifications. The coefficients and associated confidence intervals for the age–dealer interactions are graphically reported in Figure 2b. The interaction coefficients are precisely estimated, and increase monotonically until age 10 and thereafter level off and fall slightly.

Based on the estimates, we compute the *predicted dealer premium* as a price difference in dollars for each car age, and display the results in Figure 2c.[10] For all specifications, the age profile of the average dealer premium is hump-shaped and reaches its peak at age 6, at a value of between \$3,500 and \$4,000, depending on the specification. This is a large premium given that the average price of a 6-year-old dealer car is roughly \$12,000 (see Figure 1b). After age 6, the price premium declines monotonically until age twenty (less than \$1,000). Moreover, we compute the predicted price ratio of dealer sales over unmediated sales by car age and display the results in Figure 2d. The price ratio is increasing in car age until age 10, with a value of approximately 2 at that age, and then flattens and decreases slightly after age 10. It is not surprising that our estimates are noisy for older cars, because dealers sell very few old cars; see Figure 1a.

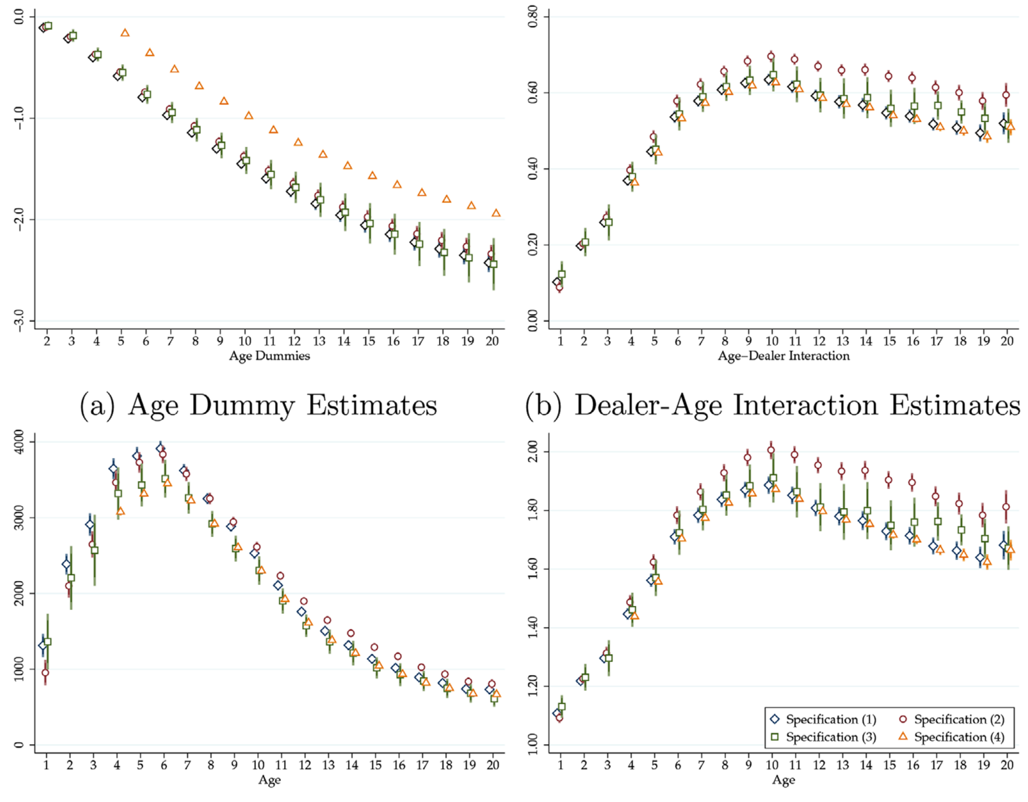
To summarize, our data suggests the following pattern of the dealer price premium.

*Fact 1.* The dealer price premium in dollar terms is positive, and it is hump-shaped with respect to car age. The dealer price premium in percentage terms is increasing in car age.

[10] Prediction in levels involves a non-linear transformation of the estimates because our dependent variable is in logs. The standard errors are adjusted accordingly. More formally, our dependent variable is $y_i = log(price_i)$. The dealer premium in levels for a given age is $exp(y|_{age,dealer}) - exp(y|_{age,direct})$ and as a ratio is $exp(y|_{age,dealer})/exp(y|_{age,direct})$.

FIGURE 2

ESTIMATION RESULTS OF DEALER PREMIUM USING THE VA DATA
[Color figure can be viewed at wileyonlinelibrary.com]




(a) Age Dummy Estimates (b) Dealer-Age Interaction Estimates

(c) Predicted Price Difference (d) Predicted Price Ratio

Note: This figure reports the point estimates with 99% confidence intervals. Different specifications refer to the different columns in Table 2.

Fact 1 is robust to the following sensitivity analysis: controlling for seller county effects, using prices in levels instead of logs, separately estimating for low-demand years (2008–2010), separately estimating for different months (due to cyclicality of inventory), separately for dealers with different size, controlling for buyer location-time effects, matching method, and using similar data from Pennsylvania (see Section 4). In addition, the dealer premium is not an average effect, but the entire distribution of dealer prices first-order stochastically dominates the distribution of unmediated market prices. All of the additional analysis can be found in online Appendix.

□ **Post-transaction resale rate and car source.** To examine the relationship between the resale rates and car source, we must be able to trace the transaction history of cars. One limitation of our VA data is that we do not observe the full VIN, and as a result we cannot follow a car's transaction history. To deal with this issue, we obtain another dataset of used car registrations that includes the full VIN from the Pennsylvania Department of Transportation (PA data). It covers all used car transactions registered from January 2014 to July 2016. The advantage of this dataset is that it includes the full VIN through which we can follow a car's post-transaction records. However, its shorter time panel makes it less ideal to test the price premium hypothesis.

**TABLE 3 Descriptive Statistics of the PA Data**

| | Dealer Sales | | Unmediated Sales | |
|---|---|---|---|---|
| No. of Initial Sales | 719,619 | (53%) | 647,645 | (47%) |
| Car age of initial purchase | 5.76 | (4.16) | 9.70 | (9.36) |
| Mileage of initial purchase | 64,392 | (52,675) | 107,978 | (67,680) |
| Price of initial purchase | $16,082 | ($9,551) | $7,395 | ($7,962) |
| **Resale Rates** | | | | |
| Resale within one quarter | 3,736 | (0.52%) | 13,755 | (2.12%) |
| Resale within two quarters | 7,325 | (1.02%) | 22,876 | (3.53%) |
| Resale within three quarters | 11,284 | (1.57%) | 31,245 | (4.82%) |
| Resale within four quarters | 15,711 | (2.18%) | 39,890 | (6.16%) |

Note: The PA data includes 1,367,264 unique cars transacted in Pennsylvania from January 2014 to July 2015. Data source: Pennsylvania Department of Transportation.

The Pennsylvania data includes 2,350,958 used car transactions registered at the Department of Transportation from January 2014 to July 2016. Among them, 53.54% were sold by dealers and the remaining 46.46% were sold directly by owners. First, we drop 48,760 transactions either because the car age is older than 20 years or because the recorded odometer is lower than 5,000 miles or higher than 300,000 miles. Second, we drop 99 transactions with the transaction price being lower than $500 or higher than $50,000. Third, we drop 26,430 transactions because they are extremely unpopular products, which we define as fewer than 30 transactions during this time period. In the end, we are left with 2,275 669 transactions in the sample, with 53% being dealer sales and 47% being unmediated sales. To study how the resale rate relates to the car source, we focus on the transactions that occurred from January 2014 to July 2015, leaving the last year as a time window of post-purchase transactions. In the end, we have 1,367,264 unique cars transacted during this period, with 719,619 cars (53%) being sold by dealers.

We define a resale as a VIN that appears multiple times in our Pennsylvania transactions dataset. Among all 1,367,264 initially transacted cars, 153,892 (11%) were resold before July 2016. Of these resales, we exclude any VIN where the second transaction was *sold* by a dealer. We do not observe private to dealer transactions, so it is likely that these are cases where the first buyer that we observe sold or traded in the car to a dealer first (we examine this behavior in Section 4). We end up with 90,911 resales between January 2014 and July 2016, where the initial seller was either a dealer or individual, the initial buyer was an individual, and the resale seller and buyer were individuals.[11]

The top panel of Table 3 reports summary statistics of our PA data, including the number of transactions, transaction price, car age, and odometer mileage separately for dealer sales and unmediated sales. The dealer sales accounts for 53% of all transactions in this data, a slightly lower than this share in our VA data. Overall, cars in the PA data are slightly newer and more expensive than those in the VA data, mainly because the transactions in the PA data occurred after those occurred in VA data. This data feature is consistent with the overall trend of the U.S. used car market during this time period. Same as in the VA sample, cars sold by PA dealers were newer and more expensive than those sold on the unmediated market. The bottom panel of Table 3 reports the share of resales within different time windows, that is, one quarter, two quarters, three quarters, and four quarters, across different car sources where the two sources are initially buying from a dealer and initially buying from a seller. Regardless of the post-transaction time windows, the resale rates of dealer cars are substantially lower than those of cars sold by sellers. For example, 0.52% of dealer cars were resold within one quarter after transaction, in contrast to 2.12% of cars sold without dealer intermediation.

[11] We also conduct our analysis with the original 153,892 resale transactions and find very similar results.

**TABLE 4 Car Source and Resale of the PA Data: FE Logit**

| | Resale Time Window | | | |
|---|---|---|---|---|
| | One Quarter | Two Quarters | Three Quarters | Four Quarters |
| Bought from Dealer | −0.761 | −0.615 | −0.532 | −0.478 |
| | (0.022) | (0.016) | (0.013) | (0.009) |
| Log Mileage | 0.238 | 0.278 | 0.283 | 0.292 |
| | (0.022) | (0.017) | (0.014) | (0.012) |

Note: The dependent variable is an indicator for post-purchase resale within the specified time window. All specifications include model–model year–trim–car age fixed effects, monthly dummies, and county indicators. Standard errors in parentheses. The data includes 1,367,264 unique cars transacted in Pennsylvania from January 2014 to July 2015. Sample selection is described in text. Source: Pennsylvania Department of Transportation.

To further understand how the likelihood of a car being resold is related to where it was bought, we estimate a Logit model with *product* (model–model year–trim–car age) fixed effects that control cars' observable characteristics, analogous to our empirical strategy for the price regression in Section 2:

$$y_i = 1\left\{\mu_i + \beta_d d_i + \boldsymbol{x}_i \boldsymbol{\beta}_{\boldsymbol{x}} + \epsilon_i > 0\right\} \tag{1}$$

where $y_i$ indicates whether car $i$ was resold within a specific time frame after transaction, $\mu_i$ are fixed effects at the model–model year–trim–car age level, $d_i$ indicates whether the car was bought from a dealer, $\boldsymbol{x}_i$ is a vector, including the log of odometer mileage when the car was bought, monthly dummies, and indicators for the buyer's county to account for local differences in selling behavior, and $\epsilon_i$ is an error term distributed i.i.d. Gumbel.

In Table 4, we report the estimation results of the Logit model for each of the four post-purchase resale time windows. Our primary coefficient of interest is the coefficient on whether a car was originally bought from a dealer ($d_i$). Our estimation results indicate that dealer cars are less likely to be resold for all four time windows we consider. Furthermore, this effect is decreasing in the number of quarters after purchase, which is intuitive if defects can usually be discovered soon after purchase.

One concern is that the buyer's purchasing decisions, and therefore outcome, may depend on unobservable characteristics that correlate with the decision to resell, potentially biasing estimates of $\hat{\beta}_d$. In other words, we are worried that $d_i$ and $\epsilon_i$ are correlated in the Logit regression described by equation (1). For example, transient individuals (e.g., short-term employees or visiting family members) who are likely to resell quickly may find it more convenient to buy from a dealer. Some individuals who buy directly from other individuals may do so as a hobby and therefore often buy and sell cars directly.

To address this potential endogeneity issue, we use a two-step control function estimation approach, following Adams, Einav, and Levin (2009)'s analysis of delinquencies on subprime car loans. To do this, we need some variable that affects a buyer's choice of whether to buy from a dealer but does not directly affect her reselling decision. We propose using dealers' inventories of cars. In particular, we compute the inventory available from all dealers in the same zip code for cars of the same body type (sedan, SUV, coupe, etc.) as the purchased product in the same week when the purchase occurred. The rationale is that greater dealer inventory could provide buyers with more options and could attract more buyers to dealers and away from the unmediated market, so it should be correlated with $d_i$. High levels of inventory may also put downward pressure on prices in local markets because of (1) tighter competition across dealers and (2) the opportunity costs associated with inventory capacity for a particular dealer lot. On the other hand,

**TABLE 5 Summary Statistics of the PA Inventory Data**

| | Mean | SD | Q25 | Median | Q75 |
|---|---|---|---|---|---|
| Dealer–Week Inventories | 55.15 | 55.63 | 19 | 41 | 75 |
| –Convertible | 2.00 | 1.59 | 1 | 1 | 2 |
| –Coupe | 3.01 | 2.56 | 1 | 2 | 4 |
| –Hatchback | 4.41 | 4.25 | 2 | 3 | 6 |
| –Minivan | 4.06 | 5.40 | 1 | 3 | 5 |
| –SUV | 21.84 | 23.52 | 7 | 16 | 31 |
| –Sedan | 24.08 | 25.53 | 8 | 17 | 32 |
| –Wagon | 2.85 | 2.14 | 1 | 2 | 4 |
| Zipcode–Style–Week Inventories | 23.28 | 62.14 | 1 | 4 | 17 |

Note: The inventory data includes 24,752 observations at the zipcode–style–week level in four areas of Pennsylvania from the 27th week of 2015 until the 8th week of 2016. Source: Cars.com.

it is unlikely that initial inventories are an important determinant of whether a buyer resells many weeks later.[12]

We obtained the dealer inventory information for transactions that occurred in four market areas from the 27th week of 2015 to the 8th week of 2016 from cars.com. Our merged dataset includes 72,538 unique used cars transacted in those areas during this period, along with their post-transaction records until July 2016.[13] The summary statistics of the sample with inventory data is in online appendix. The resale rates are similar to the rates for our baseline sample.

Table 5 displays summary statistics of inventories at the dealer level, broken down by style of car (the top panel). Dealers have roughly 55 cars on their lots on average, but there is substantial variation across dealers. There is also substantial variation across styles of cars. Sedans and SUVs are by far the most popularly offered styles of cars, which mirrors purchasing patterns. In the bottom panel of Table 5, we display summary statistics for inventories at the level of observation that we employ in our analysis, a zip code–body style–week. On average there are roughly 23 cars available for the average style in the average zip code, although this average masks large variation in inventory across styles, as can be seen in the first panel.

Our instrument relies on the assumption that current inventories do not affect the buyer's decision to resell quickly—up to 6 months after the initial purchase. To evaluate this assumption, it is necessary to understand how dealers acquire inventory. A dealer's primary source of cars is wholesale auction. Many dealers, particularly dealers with new-car franchises, rely on trade-ins as well. Trade-ins are often either re-sold at auctions to other dealers or traded to commonly owned dealers; Larsen (2014) and Murry and Schneider (2015) provide details of wholesale car markets. In Figure 3, we show that there is substantial variation in inventories across time, likely due to lumpiness and timing of auction markets and trade-ins. We break the data down by county and style of car. Each plot displays the county inventory (more aggregate variation than we use in our regressions) by style as a percentage of the inventory we observe during the first week of our data, for four counties. In some counties, inventories of different styles track each other across time, whereas in other counties this is not the case. In some instances inventories are very stable, but in other cases inventories change substantially over time.

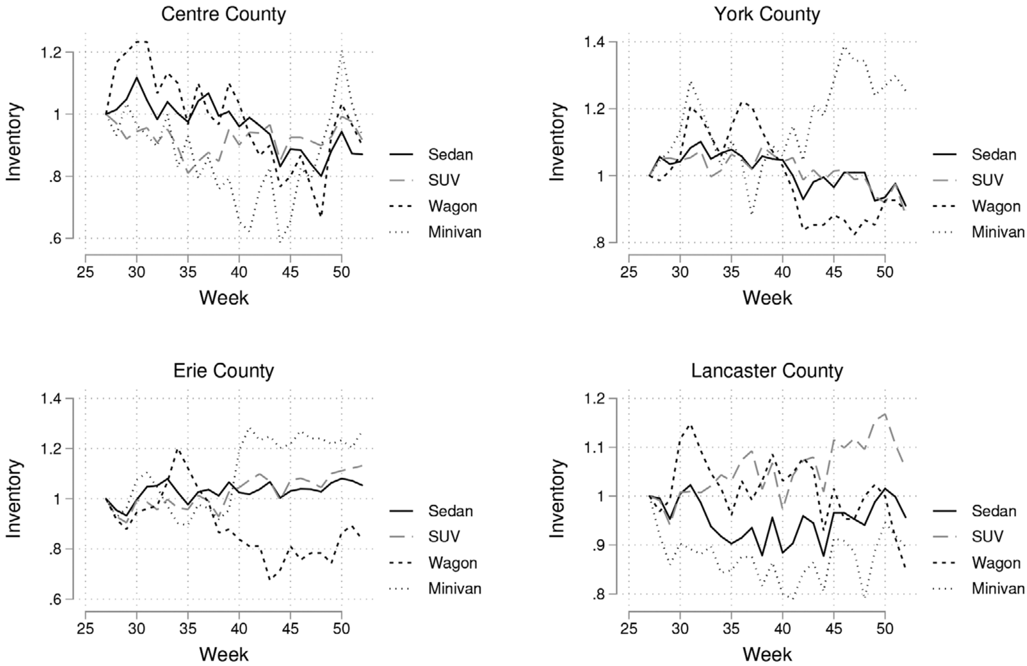
One particular story that might threaten our identification is that there may exist an aggregate shock to new car purchasers which increases the quality of the marginal car traded in. Therefore, dealers would have, simultaneously, higher inventory and better cars, and we should expect less reselling of dealer-bought cars not because of the mechanisms in our model, but due

[12] It is not our intent to separate aggregate supply and demand, as is typical when employing exclusion restrictions in estimations of market behavior. Instead, we are worried that, on the demand side, there could be individual attributes for reselling quickly that make it more likely that the original sale was from a dealer or individual.

[13] Conversations with cars.com lead us to believe that most large dealers use the platform and users typically (contractually) list their entire inventory on the platform.

FIGURE 3

COUNTY INVENTORIES BY STYLE




to the aggregate new-car shock. However, the patterns in Figure 3 do not seem consistent with this aggregate shock story, and seem more consistent with a more idiosyncratic process by which dealers acquire and manage inventory. For example, it appears that inventory is cyclical over the course of 2–3 weeks, which might be explained by patterns of auto auctions due to institutional reasons as opposed to consumer preferences.

We also instrument for dealer sales using the buyer's distance to a CarMax location. CarMax, a large national chain, typically has one of the largest inventories in a given local area and a very large virtual inventory because it can source cars from different CarMax locations. Mechanically, buyers who live near a CarMax may be more likely to purchase from a dealer just because they are likely to buy from CarMax. Also, the existence of a CarMax could force fiercer competition among dealers, driving prices down in local markets and making all dealer sales more attractive.

In the first stage, we run regressions of whether the car was originally purchased from a dealer on local dealer inventories (our excluded variable) and other variables in the resale outcome equation. The estimation results are reported in online appendix. The estimate of the coefficient before the excluded variable is positive and significant at the 10% level, which is consistent with our expectation that a used car buyer is more likely to buy from a dealer if the dealers in her neighborhood have a larger inventory of the car types she is interested in. In the second stage, we include the residuals from the first-stage regression in our Logit regression of resales.

We consider two time windows: one quarter and two quarters after transaction. The estimation results are reported in Table 6. Panel (I) reports the results of the Logit model with model–model year–trim–car age fixed effects, and panel (II) reports the results of the control function approach using the local dealers' inventory as the instrument for $d_i$. Again, cars bought from dealers are less likely to be resold shortly after purchase, with the effect being stronger for the one quarter than for two quarters. The estimates of the dealer coefficient using the control

**TABLE 6 Car Source and Resale of the PA Inventory Data: Control Function**

| | (I) FE Logit | | (II) Control Function IV: Inventory | | (III) Control Function IV: Distance to CarMax | |
|---|---|---|---|---|---|---|
| | One Quarter | Two Quarters | One Quarter | Two Quarters | One Quarter | Two Quarters |
| Bought from Dealer | −0.908 | −0.765 | −0.924 | −0.781 | −0.921 | −0.767 |
| | (0.096) | (0.070) | (0.103) | (0.075) | (0.095) | (0.070) |
| Log Mileage | 0.380 | 0.459 | 0.156 | 0.348 | 0.666 | 0.906 |
| | (0.108) | (0.082) | (0.337) | (0.216) | (0.270) | (0.209) |

Note: The dependent variable is an indicator for post-purchase resale within the specified time window. All specifications include model–model year–trim fixed effects, weekly dummies, and county dummies. We use the dealer inventory in Panel (II) and the log of the distance to the nearest CarMax in Panel (III) as the excluded variable for whether a car was bought from a dealer. Standard errors in parentheses. The sample includes 72,504 used cars transacted in four areas of Pennsylvania from the 27th week of 2015 until the 8th week of 2016. Source: Pennsylvania Department of Transportation and Cars.com.

function approach are more negative, implying a positive correlation between $d_i$ and $\epsilon_i$ in the Logit regression equation (1).

As a robustness check, we use our alternative instrument, the log of the distance between the buyer and the nearest CarMax store. The first-stage results can be found in online appendix. The estimate of the coefficient before the excluded variable is negative and significant at the 5% level, which is consistent with our expectation that a used car buyer is less likely to buy from a dealer if she is farther away from a CarMax location. The panel (III) of Table 6 reports the second-stage results. Even if we use a different exclusion restriction in the first stage, our results still suggest that dealer cars are less likely to be resold shortly after purchase. As an additional robustness check, we implement linear probability models for both the first stage and the second stage. Also, we restrict our sample to Sedan and SUV only. The results of all of these robustness analyses suggest that dealer cars are less likely to be resold shortly after purchase.

*Fact 2.* Cars purchased from dealers are less likely to be immediately resold than those purchased directly from sellers.

□ **Discussion.** Our empirical evidence leads us to conjecture that one role that dealers play in this market is to offer higher-quality products than can be obtained in the unmediated market. First, it is natural to believe that the car age affects the distribution of quality of cars and therefore the quality and price premium of the dealers. Although dealers' pretransaction service such as alleviating search frictions may contribute to the positive price premium, the value added of these service is less likely to rationalize the age pattern of the price premium. Second, the significant difference in resale rates between cars sold by dealers and cars sold directly also indicates quality differences between dealer and directly sold cars. Intuitively, the dealers' pretransaction service should have very limited impact on buyers' post-transaction decisions if the quality distributions of cars sold in the two markets (dealer and unmediated) are identical. In Section 3, we formalize a model where dealers provide high-quality products and the implications are consistent with the aforementioned empirical regularities. Following the literature on intermediaries, the model suggests two possible explanations for why dealers would find it optimal to offer higher-quality products than individual sellers: an information certification motive and an observed quality sorting motive.

## 3. Theory

■ In this section, we construct a model to rationalize the dealer quality premium. The purpose of this exercise is twofold. First, we construct a model and show that the empirical observations presented in Section 2 can be explained by the quality dynamics of cars, providing a foundation of the empirical strategy we employed. Second, we derive further testable implications based on the hypothesis of quality dynamics. We will focus on two selection mechanisms based on different sources of market frictions, the two most prevailing roles of intermediaries in the literature. After describing the basic ingredients of the model, we introduce *asymmetric* information into the model: a car's quality is privately known by the seller and the dealer. In this setting, the dealer can be viewed as a *screening device*, and obtains profits by credibly selecting high-quality cars that buyers are unable to tell from lemons. Last, we examine the model with *imperfect information* and *consumer heterogeneity*. In this setting, the dealer serves as a *sorting device* facilitating the transaction between sellers with high-quality cars and buyers with high valuations. In this case, the dealer obtains profit as it finds high-quality cars that are too costly or unlikely to be found by buyers themselves in the unmediated market.

□ **Environment.** There is a continuum of sellers, a continuum of buyers, and a monopoly dealer. Each seller owns a car. Given our modeling approach described below, we can treat each observationally equivalent car as an individual submarket in isolation.

The quality of a car is either high (H) or low (L). A car's age is $t \in [0, +\infty)$, and its quality changes over time by the following stochastic process: when new, $t = 0$, the car is of high quality. At each moment $t$, a quality shock arrives at a (failure) rate $\lambda_t$. Upon the arrival of the quality shock the car becomes low quality, $\theta_t = L$, it becomes a lemon. We assume that low quality is an absorbing state.

A seller remains passive until he receives a liquidity shock. A seller must sell his car upon the arrival of the liquidity shock.[14] The car's vintage, $t$, is publicly observable. Denote $q_t$ as the probability that a car for sale is high quality conditional on its vintage $t$. Hence, by Bayes' rule, the process of $\{q_t\}_{t\geq 0}$ must obey the following differential equation:

$$\dot{q}_t = -\lambda_t q_t, \tag{2}$$

with the initial condition $q_0 = 1$. As long as $\lambda_t \geq 0$, $q_t$ decreases in $t$. Equation (2) reflects our definition of quality and the underlying assumptions of the empirical strategy in Section 2: every car's quality is age-independent, but the average quality of cars is decreasing in car age.

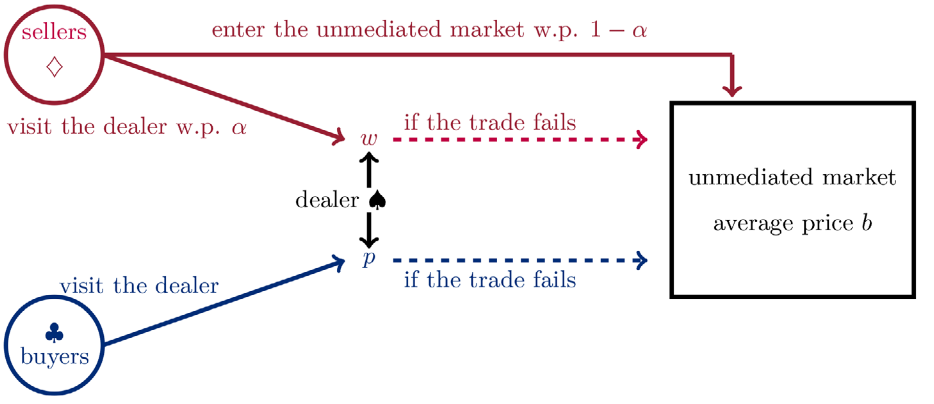
We assume the matching between a seller and the dealer is exogenous: a seller meets (or gets a price quote from) the dealer with probability $\alpha \in (0, 1)$ and goes to an unmediated market directly if he either fails to meet or does not make a transaction with the dealer. The $\alpha$ term is a reduced-form modeling device which captures the probability that a seller cannot or decides not to sell through the dealer for non-modeled reasons. We note that the seller will be indifferent between going to the dealer or to the unmediated market directly. What matters is that it ensures that some high-quality cars will be traded in the unmediated market. A seller's payoff equals the transaction price if he sells the car and zero, otherwise.

There are two types of buyers: high- and low-valuation buyers. If a buyer pays $p$ for a car of vintage $t$ whose quality is $\theta$, her payoff is $U_t^\theta - p$ if she is high valuation and it is $\phi U_t^\theta - p$ if she is low valuation, where $U_t^\theta$ represents the buyer's life time payoff of owning a $\theta$ quality car of vintage $t$ and $\phi \in (0, 1]$. A buyer is high valuation with probability $\psi_t \in (0, 1)$. In the baseline model, we assume that a buyer's valuation is her private information. We normalize $U_t^L = 0$ and let $U_t^H > 0, \forall t$. When $\phi < 1$, the Spence–Mirrlees condition holds: the high-valuation buyer values high-quality cars more than the low-valuation buyers. We assume that $\dot{U}_t^H \leq 0$, and, $\lim_{t\to\infty} U_t^H = 0$, to capture the *depreciation effect*. That is, as the car ages, the marginal benefit of owning a high-quality car rather than a low-quality one is falling and eventually vanishes, implying that the effect of quality on price is also falling as the car ages.

[14] We abuse the term of a liquidity shock to capture exogenous reasons for which the seller has to sell his car. Examples include the need to buy a new car, moving to other countries (states), etc.

FIGURE 4

ILLUSTRATION OF THE ORDER OF MOVES
[Color figure can be viewed at wileyonlinelibrary.com]




The red ◇ represents a cohort $t$ seller whose car quality $\theta_t$ is either $L$ or $H$, the blue ♣ represents a buyer whose willingness to pay is either $U_t^{\theta_t}$ or $\phi U_t^{\theta_t}$, and the black ♠ represents the dealer who posts a purchase price $w$ and the selling price $p$.

A buyer purchases from either the seller or a dealer. In either case, we assume the buyers have no bargaining power. When a buyer meets a seller or dealer, the owner of the car makes a take-it-or-leave-it offer. A buyer does not observe the price offers made to other buyers. For simplicity, we assume that every buyer automatically visits the dealer first. If a buyer fails to purchase a car from the dealer, she goes to the unmediated market.

The dealer has monopoly power. He makes a private take-it-or-leave-it offer to each seller and buyer who visits him. The dealer's payoff equals the total revenue from selling cars, minus the total cost of purchasing cars, and reputation cost due to selling lemons. We let $p$ be the selling price to a buyer; $w$ is the purchasing price to a seller. We let $k > 0$ be the dealer's disutility due to selling a low-quality car. It can be justified as a negative net operational cost, a reputation loss, or a monetary loss due to the requirement of a warranty.

Although the quality of each car evolves over time, no trade can occur before the arrival of the liquidity shock. Thus, we treat the arrival time $t$ as a parameter and analyze the strategic interaction upon the arrival of the liquidity shock at time $t$. For simplicity, at each $t$, we assume that the measure of active sellers and buyers is equal and normalize it to one. Thus, we examine each cohort of cars in isolation.

The order of moves of cohort $t$ game, visualized in Figure 4, is given as follows: (i) Nature decides whether a seller meets a dealer (with probability $\alpha$). If a seller meets a dealer, the dealer makes a take-it-or-leave-it purchasing offer, $w$, to the seller. Then the seller decides between accepting the offer and rejecting it and going to the unmediated market. (ii) Buyers meet the dealer. The dealer makes a take-it-or-leave-it selling offer, $p$, to each buyer. Each buyer decides between accepting the offer and rejecting it and going to the unmediated market. (iii) In the unmediated market, sellers and buyers who fail to trade with the dealer randomly match pairwise, and the seller makes a take-it-or-leave-it offer. The average transaction price is denoted by $b$.

□ **Selection based on asymmetric information.** In this section, we assume that buyers are *homogenous*, $\phi = 1$ and the quality of the car $\theta_t$ is privately observed by the seller. We focus on the role of dealer as an information intermediary to deal with the information asymmetry. For simplicity, if a seller visits the dealer, the dealer perfectly observes the quality of the car $\theta_t$ and decides whether to purchase it and at what price. We assume that $k > U_0^H$ so that a dealer would not want to sell a lemon of any vintage. A buyer's prior belief that the car is of high quality is $q_t$. When a buyer and a seller meet in the unmediated market, the buyer observes neither the quality of the car nor whether the seller has visited a dealer. Notice that the information asymmetry between the seller and buyers is developing over time: as the car ages, the public prior belief declines, with as $t \to \infty$, $q_t \to 0$; see Hwang (2018) for a more detailed discussion of developing asymmetric information.

We analyze players' incentives via backward induction. We begin with the transaction in the unmediated market. Since $\theta_t$ is unobservable, a buyer's willingness to pay is $b_t = \hat{q}_t U_t^H$ where $\hat{q}_t$ denotes the equilibrium posterior belief conditional on the seller going to the unmediated market. We focus on the strategy profile where the seller's offer has no signaling effect, so the seller's optimal price is $b_t$, and the buyer accepts it for sure.[15] The seller rationally anticipates his payoff is $b_t$ if he goes to the unmediated market, so he accepts (or rejects) the dealer's offer for sure if it is strictly higher (or lower) than $b_t$, and in equilibrium, the seller will accept the dealer's offer of $b_t$ with probability 1. Notice that $\hat{q}_t > 0, \forall t$ because $\alpha < 1$.

Now, we turn to the dealer's problem. A buyer's willingness to pay for a dealer's car is $\tilde{q}_t U_t^H$ where $\tilde{q}_t$ denotes his equilibrium posterior belief conditional on the car being traded through the dealer. Since $k > U_0^H$ and $\dot{U}_t^H \leq 0$, it is never optimal for the dealer to trade a lemon. Thus, if there is any trade in the equilibrium, the dealer purchases from the seller only if $\theta_t = H$, and the buyers' willingness to pay is $U_t^H$ for the dealer's car. In equilibrium, buyers who are indifferent between accepting and rejecting the dealer's offer will mix to balance the dealer's supply and the buyers' demand. As a result, a high-quality car is traded in the unmediated market only if the seller fails to find the dealer; and thus in the equilibrium,

$$b_t = \frac{(1-\alpha)q_t}{1-\alpha q_t} U_t^H. \tag{3}$$

The numerator is the measure of high-quality cars directly sold in the unmediated market, and the denominator is the measure of all cars sold directly to buyers: those that never go to the dealer, $(1-\alpha)$, plus those that go to the dealer but are lemons which the dealer does not buy, $\alpha(1-q_t)$. To maximize his profit, the dealer makes a *minimum winning offer* $w_t = b_t$ for high-quality cars and a *losing offer* $w < b_t$ for low-quality cars. The former is the lowest offer that will be accepted by a high-quality seller, whereas the latter will be declined by a low-quality seller and results in zero payoff to the dealer. Formally,

*Proposition 1.* For any $t$, there is an equilibrium in which

(1) A seller makes a take-it-or-leave-it price $b_t$ in the unmediated market. If the seller visits the dealer, he accepts the dealer's offer only if it is at least as large as $b_t$.
(2) The dealer makes a losing offer when $\theta_t = L$ and a minimum winning offer $w_t = b_t$ when $\theta_t = H$. The dealer sells cars at price $p_t = U_t^H$.
(3) Every buyer breaks even: in the unmediated market, a buyer accepts the seller's offer if and only if the price is not higher than $b_t$ satisfying (3) in the unmediated market, and a buyer rejects the dealer's offer if the price is higher than $U_t^H$. He accepts it for sure if the price is strictly lower than $U_t^H$, accepts the offer with probability $\alpha q_t$ if the price equals $U_t^H$.

In the equilibrium, the dealer trades with the seller only if $\theta_t = H$, causing an *adverse selection* effect on the set of the sellers going to the unmediated market. Accordingly, the buyers will lower their belief of the quality of cars on the unmediated market and thus their maximal price that they are willing to accept from a seller. The average quality of the cars traded through the dealer is $U_t^H$, which is higher than that of unmediated sales, $\frac{(1-\alpha)q_t}{1-\alpha q_t} U_t^H$. The difference in the quality of cars traded through the dealer and those traded in the unmediated market reflects two

[15] Buyer beliefs off-equilibrium path that assume any different offer comes from a low-quality seller are sufficient for this.

effects, one direct and one indirect. First, the dealer has a better technology to screen a high-quality car from a low-quality car and thus he has an informational advantage. Second, because the dealer only purchases high-quality cars, the dealer's information advantage generates an adverse selection effect: it increases the proportion of low-quality cars in the unmediated market, which further enlarges the quality difference between the dealer's supply and the supply on the unmediated market.

Fixing the car's vintage and other observable characteristics, we call the difference in the transaction price at the dealership and the unmediated market the dealer *price premium*. The dealer's price premium varies as the car ages. Although both the dealer price, $p_t = U_t^H$, and the unmediated market price, $b_t = \frac{(1-\alpha)q_t}{1-\alpha q_t} U_t^H$, are decreasing in $t$, the driving forces for the declining price are different. The dealer's price declines simply because of car depreciation ($\dot{U}_t^H \leq 0$). On the other hand, the price of an unmediated transaction is decreasing because of car depreciation and it is also more likely a lemon ($\dot{q}_t < 0$). We now show that the model's implications on the price premium are consistent with our empirical results in Section 2.

First, we examine the age effect on the dealer's price premium in dollar terms:

$$p_t - b_t = \frac{1 - q_t}{1 - \alpha q_t} U_t^H. \tag{4}$$

To investigate the age effect, we take the derivative of (4) with respect to $t$ and obtain

$$\underbrace{-\frac{(1-\alpha)}{(1-\alpha q_t)^2} U_t^H \dot{q}_t}_{(+)} + \underbrace{\frac{1 - q_t}{1 - \alpha q_t} \dot{U}_t^H}_{(-)}. \tag{5}$$

The total age effect can be decomposed into two parts. First, it affects the dealer's value as an information intermediary. That is, it decreases the public prior belief $q_t$ and thus the posterior belief of the buyers in the unmediated market, lowering the unmediated market price. Consequently, it increases the dealer's premium. This is captured by the first term of formula (5). Second, it decreases the buyer's willingness to pay for a high-quality good, which is captured by the second term of formula (5). This is the standard depreciation effect. In general, the total effect of age on the premium is *non-monotonic*.

When $t = 0$, $q_t = 1$, so the second effect does not appear. Clearly, the price premium in (4) is strictly positive for $q_t < 1$, so the price premium in dollars is positive and initially increasing for small $t$. On the other hand, for very old cars, as $t \to \infty$, $U_t^H$ goes to zero, and so does the price premium according to equation (4). Therefore, the price premium must eventually fall.

Second, one can also formalize the dealer's price premium over unmediated sales in percentage terms:

$$\frac{p_t}{b_t} = \frac{1/q_t - \alpha}{1 - \alpha}. \tag{6}$$

When one takes the ratio between the dealer transaction price and unmediated transaction price, the depreciation effect, $U_t^H$, drops out, and one can isolate the age effect through the change in $q_t$, that is, the change in the dealer's value of alleviating asymmetric information. Clearly, the formula in (6) is increasing in $t$. Notice that the depreciation effect drops out in the ratio is the consequence of the separability of the depreciation effect ($\dot{U}_t < 0$) and the expected quality dynamics effect ($\dot{q}_t < 0$), which depends on neither the normalization $U_t^L = 0$ nor the binary type assumption. What matters is that two effects determine the buyers' willingness to pay as a product.[16]

[16] Alternatively, one can assume the type of car $\theta \in [1, \bar{\theta}] \subset \mathbb{R}$ and the buyer's value of a $\theta$-type car is given by $\theta U_t$ where $\dot{U}_t < 0$ and $\lim_{t\to\infty} U_t = 0$. The term $U_t$ captures the separable depreciation effect. Then the $U_t$ still drops out in the price premium in percentage terms.

Formally, we summarize our first empirical implication, which rationalizes Fact 1, as follows.

*Implication 1.* The dealer's price premium in dollar terms formulated in (4) is positive for all car ages and is non-monotone in the car's age. For recent vintages it increases, and for sufficiently old cars it decreases: it is hump-shaped. The dealer's price premium over unmediated sales in percentage terms formulated in (6) is greater than one for all car ages and is increasing in the car's age.

Recall the classic logic of Akerlof (1970): asymmetric information causes cars that are observably identical to buyers to sell for the same price even though they may actually be of different qualities. Hence, owners of unobservably high-quality cars will sell them less often because the seller's reservation prices are higher. Our theoretical analysis predicts that dealer cars are of higher unobserved quality. Therefore, we should expect that buyers of dealer cars are less likely to resell their cars because their cars are of higher average quality.

We now extend our base model by allowing post-transaction resale. Recall that in stage 3 of our base model, a buyer immediately learns the quality of the car. We add a subsequent resale stage. At this stage, a buyer receives a liquidity shock with probability $\delta \in (0, 1)$ so that he has to sell his car in a separated resale market. We also allow a buyer to sell his car even if he does not experience a liquidity shock. The resale market observes the car's vintage and that the car has been sold to the current owner recently, but can neither tell a buyer's motive for trying to sell the car nor tell whether the car was purchased from a dealer or directly from a seller. For simplicity, we do not explicitly model the demand and the transaction process of the resale market, and we assume the resale market is competitive and price equals the rational expected value of the quality of cars. As $\delta > 0$, a high-quality car is resold with a positive probability, so the resale price $R_t > 0$. On the other hand, some low-quality cars will be resold too, so $R_t < U_t^H$. Therefore, a high-quality car owner will resell his car only if he receives a liquidity shock, whereas a low-quality car owner will always resell his car.

If a buyer purchased the car from a dealer, he will resell a high-quality car with probability $\delta$. In contrast, if a buyer purchased the car from a seller directly, he will resell a car if either the liquidity shock arrives or the car is a lemon. His resale rate in this case is given by

$$\frac{(1-\alpha)q_t\delta + (1-q_t)}{1-\alpha q_t}. \tag{7}$$

The numerator consists of sellers who sell their high-quality cars directly to buyers who have a liquidity shock plus the measure of buyers who will sell their low-quality cars that buyers want to sell, $(1-q_t)$. The denominator is the measure of all cars sold directly to buyers: those that never go to the dealer, $(1-\alpha)$, plus those that go to the dealer but are lemons which the dealer does not buy, $\alpha(1-q_t)$. Clearly, a car bought directly from a seller has a resale rate greater than $\delta$. Therefore, we derive another testable implication:

*Implication 2.* A buyer is less likely to resell his car if it was purchased from a dealer.

This implication is consistent with Fact 2. Simple algebra shows that when the probability of liquidity shock is sufficiently small, the prediction regarding the dynamics of price premium in Implication 1 remains.

□ **Selection based on buyers' heterogeneity.** In this section, we propose a selection theory based on consumer heterogeneity instead of asymmetric information to rationalize the empirical evidence in Section 2. Specifically, we assume that (1) the seller's type (or his car's quality) $\theta_t$ is observed by all players but not by the econometricians, and (2) the buyer's valuation is heterogenous, that is, $\phi < 1$, so the high-valuation consumers' willingness to pay for a high-quality car, $U_t^H$, is strictly higher than the low-valuation consumers' willingness to pay $\phi U_t^H$; whereas both types of buyers' willingness to pay for a low-quality car remains zero. Furthermore, to avoid rationing and simplify the analysis, we assume that $\psi_t = \alpha q_t$. That is, the maximum quantity of high-quality cars being sold through the dealer is equal to the total measure of high-valuation buyers.

In this model, the dealer's role is to facilitate the assortative matching between cars and buyers. Since buyers value high-quality cars differently, the dealer can selectively attract high-valuation buyers and sellers with high-quality cars through his pricing because he improves the average matching efficiency. The dealer's price premium reflects not only high quality of cars but the additional matching surplus being created. In this case, the dealer cannot sell a low-quality car as a high-quality car, so any positive $k$ will prevent him from trading low-quality cars. The main result is as follows.

*Proposition 2.* There exists an equilibrium where

(1) a seller who visits the dealer accepts any purchase offer $w \geq \phi U_t^H$ if $\theta_t = H$ and any offer $w \geq 0$ if $\theta_t = L$. In the unmediated market, the seller charges $\phi U_t$ for high-quality cars and 0 for low-quality cars;

(2) the dealer only purchases high-quality cars at a price

$$w_t = \phi U_t^H \tag{8}$$

and the selling price is

$$p_t = U_t\left[1 - (1-\phi)\frac{(1-\alpha)q_t}{1-\alpha q_t}\right]; \tag{9}$$

(3) a high-valuation buyer purchases from the dealer if and only if $\theta_t = H$ and

$$p_t \leq U_t\left[1 - (1-\phi)\frac{(1-\alpha)q_t}{1-\alpha q_t}\right],$$

a low-valuation buyer purchases from the dealer if and only if either $\theta_t = H$ and $p_t \leq \phi U_t$ or $\theta_t = L$ and $p_t \leq 0$, and in the unmediated market, the buyer purchases if the price is no higher than $\phi U_t$ for a high-quality car and 0 for a low-quality car.

To see that this is an equilibrium, let us examine each player's incentives. First, suppose a high-valuation buyer is deciding whether to purchase from a dealer at a price $p$. If he declines the offer and goes to the unmediated market, sellers treat him as a low-valuation buyer. With probability $\frac{(1-\alpha)q_t}{1-\alpha q_t}$, he meets and purchases from a seller with a high-quality car at a price $\phi U_t^H$, and with the complementary probability, he meets and purchases from a seller with a low-quality car at a price 0. In this case, his expected payoff is $(1-\phi)U_t^H$. If, instead, he purchases from the dealer at a price $p$, his payoff is $U_t^H - p$. Therefore, his willingness to pay for a high-quality car is given by the price in equation (9). On the other hand, a low-valuation buyer finds it strictly suboptimal to purchase a high-quality car from the dealer at a price in equation (9) because

$$\phi U_t^H - p_t = U_t^H(1-\phi)\left[\frac{(1-\alpha)q_t}{1-\alpha q_t} - 1\right] < 0, \forall t.$$

Second, the dealer has no incentive to buy and sell low-quality cars due to the reputation cost; he can purchase every high-quality car at price $w_t$ defined in equation (8) and sell it to high-valuation buyers at price $p_t > w_t$.

Third, a seller with a high-quality cars has no incentive to decline the dealer's offer. This is because he anticipates all high-valuation buyers will go to the dealer and prefers $w_t$ to what he can get in the unmediated market. Finally, it is easy to see that a low-quality seller will only sell in the unmediated market. Hence, the above strategy profile is an equilibrium.

In the unmediated market, both high-quality cars and low-quality cars are traded; the *average price* is therefore given by $b_t = \frac{(1-\alpha)q_t}{1-\alpha q_t}\phi U_t^H$, and we can compute the dealer's price premium in difference terms, $p_t - b_t = U_t^H[\frac{1-q_t}{1-\alpha q_t}]$, which is identical to the one in equation (4), so the age effect on the price premium in difference remains. Similarly, one can compute the dealer's price premium in percentage terms, which is given by $p_t/b_t = \frac{1/q_t-\alpha-(1-\alpha)(1-\phi)}{\phi(1-\alpha)}$. Although the previous formula differs from the one in equation (1), the age effect on price premium remains qualitatively the same. Taking derivative with respect to $t$ yields $\frac{\dot{q}_t}{\phi q_t^2(1-\alpha)} < 0$. Therefore, as the car age increases, the dealer's price premium in percentage term increases, which is consistent with Implication 1, as was the case in the model with asymmetric information.

Similarly, one can extend the benchmark game by allowing reselling in a separated market. In this case, reselling is not driven by adverse selection. The motive for reselling is due to the inefficient allocation in the unmediated market: low-valuation buyers may still get high-quality cars with positive probability. By our assumption, there is positive gain from trade between high-valuation buyers who participate in the resale market and low-valuation buyers who purchased high-quality cars. Assume that the resale rate, driven by liquidity motive, is $\delta$, then if a car is purchased from the unmediated market, the resell rate is

$$\frac{(1-\alpha)q_t + (1-q_t)\delta}{1-\alpha q_t}. \tag{10}$$

The intuition behind the previous formula is as follows. In total, the measure of cars being traded in the unmediated market is $1-\alpha q_t$. In the equilibrium, they are all purchased by low-valuation buyers. If a low-valuation buyer purchased a high-quality car, he will sell it for sure to other high-valuation agents; otherwise, he will sell it with probability $\delta$. Obviously the resale rate of cars being sold in the unmediated market is higher than $\delta$. On the other hand, only high-valuation buyers purchase from the dealer. There is no gain from trade of resale, so the resale only occurs when the buyers receive liquidity shocks. As a result, the resale rate of cars sold through the dealer is $\delta$.

In sum, when the dealer's role is to facilitate sorting, the relation between the resale rate and car source is consistent with Implication 2. As this extension is very straightforward and in the same spirit of the asymmetric information story, the analysis is omitted.

□ **Discussion.** Now, we discuss the key model features necessary to derive the testable implications in this section. Four ingredients are important for our theoretical analysis. First, for each car age, the quality of cars are heterogeneous, and as cars age, both the average quality and a consumer's marginal willingness to pay for a high-quality car relative to low-quality one falls; this is embodied in equation (2) and the pricing equations. Second, there exist significant frictions due to either incomplete (asymmetric) information or imperfect information (search frictions), making it very costly for buyers to costlessly identify high-quality cars from low-quality ones; buyers cannot either observe a car's true quality before purchase in the asymmetric information setting or the buyer cannot observe the car's quality before seeing it. Third, dealers are more able to at overcoming information frictions and can credibly select high-quality cars which they sell to buyers; the dealer always perfectly observes the car quality and suffers disutility $k$ by selling low-quality cars, ensuring he never trades low-quality cars. Fourth, the transaction price reflects the buyers' expected quality of the car: buyers are rational; as embodied in the pricing and resale equations and have zero bargaining power. In what follows, we briefly discuss how our mechanism is complemented by some other institutional details such as the use of warranty, dealers' reputation mechanism, the wholesale auction of used-car, and other features.

Our analysis relies on credible selection of high-quality cars through dealers. For a dealer to work as an honest gatekeeper, he must value the quality of the cars that he trades. In our model, we directly assume that selling a lemon will cost $k$ dollars of expected profit for the dealer. This assumption captures the fact that dealers are less myopic than sellers. This can be justified in two ways. First, the disutility of selling lemon can also be justified by the use of warranty as in Biglaiser (1993) and Biglaiser and Friedman (1999). For example, CarMax offers 15-days warranties. Second, unlike a seller who only sells one car, a dealer cares about his long-term reputation (see Biglaiser and Friedman (1994)).

In the model, the dealer "refuses" to purchase low-quality cars and only sells high-quality cars to consumers. This reduced-form modeling device enables the dealer to only deliver high-quality cars from sellers to buyers. In reality, this mechanism can be implemented in many other ways. For example, low-quality car owners anticipate unattractive offers from the dealers and therefore choose to visit the dealers with a lower probability. Also, the dealers can purchase low-quality cars at a sufficiently low price and sell them to other dealers in wholesale used-auto auctions rather than to final consumers. See a more detailed discussion on the wholesale automobile auctions in Genesove (1993) and Larsen (2014). Also, a dealer may repair low-quality cars and make them high quality. Without data on dealers' purchase prices and repair decisions, it is impossible to distinguish whether the quality premium results from careful quality selection or repairing by the dealer. Fortunately, the empirical implications we derived do not depend on the selection mechanism.

## 4. Asymmetric information versus consumer heterogeneity

■ The fundamental difference between our two theories in Section 3 lies in the assumption of consumers' information sets. In the first theory, the car quality is privately observed by sellers but not buyers, and the dealer servers as a screening device that alleviates the inefficiency caused by asymmetric information. The second theory assumes that car quality is public information. Some buyers' willingness to pay for high-quality cars is higher, but it is costly for them to find high-quality cars in the unmediated market. The dealer facilitates assortative matching between heterogeneous buyers and sellers. In this section, we examine the extent to which one can empirically distinguish between the asymmetric information theory (based on unobserved quality) and the sorting theory (based on observed quality).

□ **Resale rate and car age.** We now propose distinguishing the two theories by examining the car age effect on the resale rate. By the specification of the quality process, as the car age increases, the proportion of low-quality cars increases, that is, $\dot{q}_t < 0$. In the asymmetric information theory, a buyer resells his car because either he realizes that his recent purchase is a lemon or he receives a liquidity shock. Since we assume that a liquidity shock is independent of a car's age, then the probability that a car is resold due to its being a low-quality car increases as the car gets older. To isolate these effects as much as possible, we examine the age effect on the gap between the resale rates of cars purchased from sellers and cars purchased from dealers, and we focus on those rates within one or two quarters after the initial transaction. Specifically, the resale rate is given by equation (7) for cars purchased from the unmediated market, and it is $\delta$ for cars purchased from the dealer. Simple algebra implies that the resale rate gap between the unmediated market cars and dealer's cars is

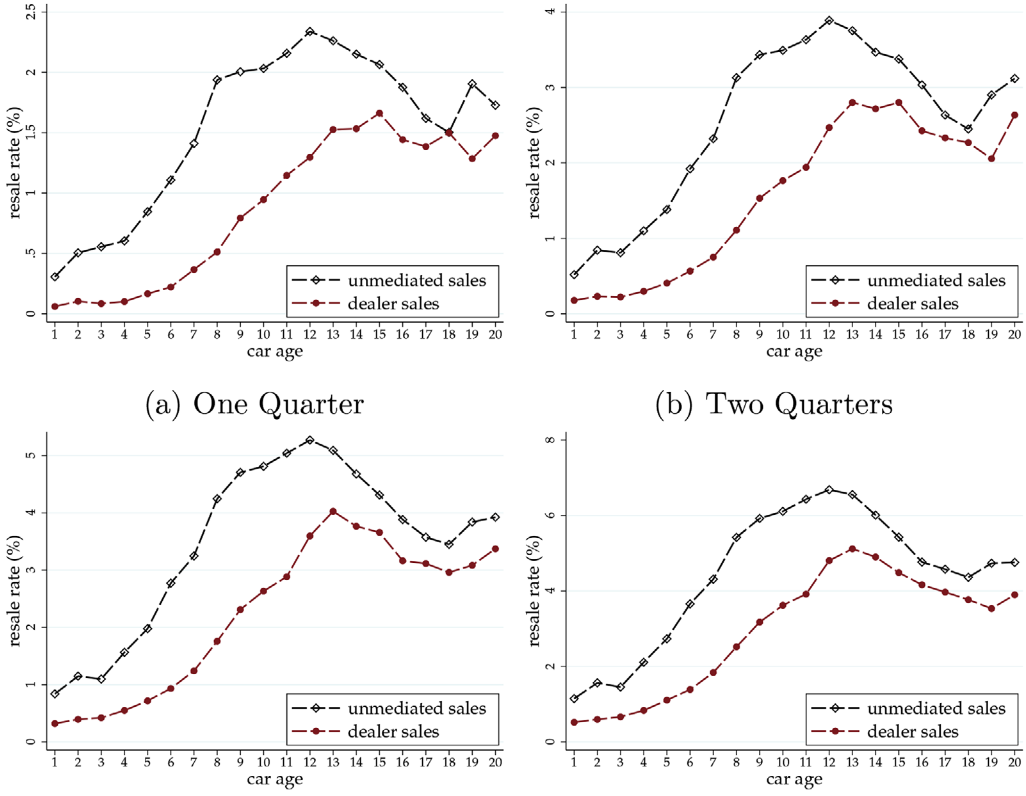
$$1 - \delta - (1 - \alpha)\frac{1 - \delta}{\frac{1}{q_t} - \alpha}, \tag{11}$$

which is decreasing in $q_t$. Therefore, the asymmetric information theory predicts that the resale rate gap is increasing in car age.

On the other hand, in the sorting theory, resale takes place to correct an allocation inefficiency in the initial transaction: a low-valuation buyer who purchased a high-quality car finds it profitable to sell it to someone who has a higher valuation. Thus, high-quality cars are resold. As the car age increases, the proportion of high-quality cars becomes smaller, and so does the resale rate. Specifically, the resale rate is given by equation (10) for cars purchased from the

FIGURE 5

RESALE RATES PREDICTED BY FE LOGIT USING PA DATA [Color figure can be viewed at wileyonlinelibrary.com]




(a) One Quarter (b) Two Quarters (c) Three Quarters (d) Four Quarters

unmediated market, and it remains $\delta$ for cars purchased from the dealer. The resale rate gap between the unmediated market cars and the dealer's cars is

$$\frac{(1-\alpha)(1-\delta)}{\frac{1}{q_t}-\alpha}, \tag{12}$$

which is increasing in $q_t$. Therefore, the sorting theory predicts that the resale rate gap is decreasing in car age. Thus, one can distinguish the two theories by examining the age effect on the resale rate gap between the unmediated market cars and dealer cars. Below, we will use the PA data that allows us to keep track of the transaction history of cars to examine the age patterns of the resale rates. The data is described in Section 2.

We re-estimate equation (1) but add interactions of the car source and car age dummies. Based on the estimates, we compute the predicted resale rates by car age separately for cars purchased from the unmediated market and cars purchased from dealers and plot them in Figure 5. First, unmediated sales have higher resale rates for cars of all ages. Second, both resale rates of unmediated sales and dealer sales are hump-shaped in the age of the car. Both rates increase in car age until age 12 and then decline. Since the sample includes very few observations for really old cars and hence the pattern is very noisy after age 18. Last, the difference of resale rate between unmediated sales and dealer sales is hump-shaped in car age, peaking at age 9. Graphically, the gap between the black and red lines becomes wider as the age increase until age 9 and then narrows as cars become very old. In addition, the ratio of the resale rate of directly sold cars and that of dealer cars is the greatest for one-quarter resales.

Our conclusion is that the information theory seems to dominate for younger cars (the diverging resale rates for cars younger than 9 years old) and the sorting theory dominates for older cars (the converging and decreasing resale rates). This result can be interpreted as a dynamic decomposition of the results in Peterson and Schneider (2014). They view a car as an assemblage of parts, some with asymmetric information, and others without, and found evidence of adverse selection and consumer sorting. Our result not only confirms their finding, but also identifies which sets of cars the informational asymmetry are most important.

Again, to account for potential endogeneity of the initial sale, we use the local dealer inventory as the instrument. We focus on the sample that has the dealer inventory information; see data description in Section 2. Recall that the sample size with the inventory observations is relatively small. To ensure the reliability of the estimation of the age effect, we coarsen the age categories and divide cars into the following age groups: (i) very young cars ages 1–3, (ii) young cars ages 4–6, (iii) medium-age cars ages 7–10, (iv) old cars ages 11–15, and (v) very old cars beyond 15 years. Table 7 reports the number of sales, percentages of resales within one quarter and two quarters across the five age groups.

To examine the age effect, we run a regression of whether a car was sold within a time period (one quarter and two quarters) or whether the car was bought from a seller, interactions of seller and car age category dummies, the log of the mileage, weekly dummies, and county dummies. In this regression, we further control for car characteristics using fixed effects at the car model and trim level (as in all of our previous analysis). The first and third columns of Table 8 report the estimates of the interaction terms that capture the resale differences of unmediated sales and dealer sales. The age–source interactions paint a similar picture as in Figure 5—the difference of resale rate between direct and dealer sales is small and insignificant for very young cars, becomes pronounced for medium-age cars, and becomes insignificant again for very old cars.

To address the endogeneity problem, we instrument the car source with the dealer inventory at the zip code–style–week level, as in Section 2. In the first stage, we run a regression of whether a car was bought from a seller on the local dealer inventory, age category dummies, log of mileage, weekly dummies, and county dummies, controlling for the product fixed effects at the car model and trim level. The estimate of the excluded variable coefficient is negative and significant at the 10% level, consistent with our expectation that a larger dealer inventory discourages used car buyers to buy from sellers. The second-stage estimates of the interaction terms are reported in the second and last columns of Table 8. The estimates of the coefficient before the indicator of very young cars (1–3 years) are not significant at the 10% significance level. The estimates for very old cars are noisy, because there are both very few dealer and unmediated sales. The estimates for the other three age categories are significantly positive, with the estimates for the medium-aged cars being the largest. Formally,

*Fact 3.* The difference of resale rate between unmediated sales and dealer sales becomes wider as cars become old, but then narrows as cars become very old.

Fact 3 suggests that alleviating asymmetric information is the dominant role of dealers for most used cars, although dealers primarily promote sorting efficiencies as cars become very old.

□ **Resale price.** Another way to empirically distinguish the two theories is to examine the transaction prices of the resales we observe in our data. The asymmetric information story suggests that quick-resale cars are more likely to be lemons than when they were purchased, so the equilibrium resale price will be lower than the price at which they were purchased; whereas with sorting, resale cars are more likely to be of high quality, so the equilibrium resale price will be higher.

**TABLE 7 Resale and Car Age: PA Inventory Data**

| | Dealer Sales | | | Unmediated Sales | | |
|---|---|---|---|---|---|---|
| | No. of Sales | One quarter resales (%) | Two quarter resales(%) | No. of Sales | One quarter resales (%) | Two quarter resales (%) |
| (i) Very Young | 18,280 | 0.09 | 0.19 | 4,357 | 0.55 | 0.96 |
| (ii) Young | 8,699 | 0.11 | 0.32 | 3,283 | 1.07 | 1.74 |
| (iii) Medium | 9,843 | 1.64 | 3.01 | 9,145 | 2.33 | 3.94 |
| (iv) Old | 5,550 | 2.49 | 4.44 | 10,680 | 3.16 | 5.18 |
| (v) Very Old | 644 | 1.86 | 3.88 | 2,060 | 2.96 | 5.39 |
| All age | 43,016 | 0.43 | 0.86 | 29,522 | 2.27 | 3.80 |

Note: The PA inventory data includes 72,538 used car transactions registered in four areas of Pennsylvania from the 27th week of 2015 until the 8th week of 2016. Sample selection is described in Section 2. Source: Pennsylvania Department of Transportation and Cars.com.

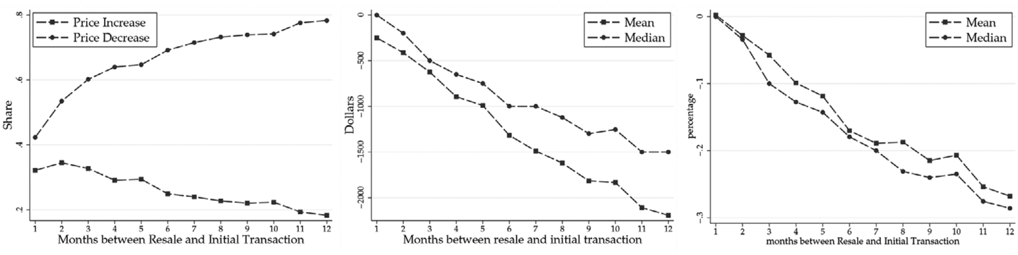
**TABLE 8 Resale Difference between Unmediated and Mediated Sales: PA Inventory Data**

| | One Quarter Resale | | Two Quarter Resale | |
|---|---|---|---|---|
| | OLS | IV | OLS | IV |
| (i) Very Young | 0.003 | −0.075 | 0.006 | 0.012 |
| | (0.002) | (0.137) | (0.002) | (0.179) |
| (ii) Young | 0.005 | 0.038 | 0.005 | 0.060 |
| | (0.003) | (0.014) | (0.004) | (0.019) |
| (iii) Medium | 0.012 | 0.075 | 0.017 | 0.120 |
| | (0.002) | (0.014) | (0.003) | (0.018) |
| (iv) Old | 0.009 | 0.065 | 0.011 | 0.096 |
| | (0.003) | (0.016) | (0.007) | (0.021) |
| (v) Very Old | 0.007 | 0.048 | 0.006 | 0.046 |
| | (0.005) | (0.030) | (0.001) | (0.040) |

Note: The dependent variable is an indicator for post-purchase resale within the specified time window. All specifications include log of mileage, model-trim fixed effects, weekly dummies, and county dummies. In the IV columns, we use dealer inventory as the excluded variable for whether a car was bought from a dealer. Standard errors in parentheses. The PA inventory data includes 72,538 used car transactions registered in four areas of Pennsylvania from the 27th week of 2015 until the 8th week of 2016. Source: Pennsylvania Department of Transportation and Cars.com.

FIGURE 6

RESALE PRICE AND INITIAL TRANSACTION PRICE: PA DATA




(a) Price Change Direction (b) Price Change in Dollars (c) Price Change in Percentage

Note: The PA data includes all resales that occurred before July 2016 of all used car transactions registered in Pennsylvania from January 2014 to July 2015.

As we described in Section 2, more than 1.4 million used car transactions were registered in Pennsylvania from January 2014 to July 2015, among which 11% were resold before July 2016. To get a sensible sample to examine the change in the resale price relative to the initial transaction price, we exclude obvious price outliers.[17] For each resale, we calculate the difference between the resale price and its initial transaction price.

Figure 6a presents the share of resales with price increase and the share of resales with price decrease by how long the resale occurred after the initial transaction, measured in months. When a resale occurs only 1 month after the initial transaction, it is equally likely that the resale price is higher or lower than the initial price. However, as the duration between the resale and the initial transaction grows, the resale price is more and more likely to be lower than the initial transaction price. In particular, when the duration is 12 months, the resale price is almost always lower than the initial price.

In Figures 6b and 6c, we present the mean and median price changes in dollars and in percentage terms. When resales occur only 1 month after the initial transactions, both the mean and the median of the gap between the resale price and the initial price are almost zero. As the

[17] If a transaction price is four times higher or one quarter lower than the predicted price from a hedonic regression, we consider it as an outlier.

duration gets longer, the difference between the resale price and the initial price falls. After 2 months, the median decrease is roughly 4%, and after 3 months, the median decrease is roughly 6.5%.

*Fact 4.* The resale price of a car can be either higher or lower than its initial transaction price, but it is more likely lower. Over time, the proportion of cars with a negative price change increases. Both the average and median of the price changes are negative and decrease over time.

The fact that prices can either rise or fall relative to the original transaction price suggests that both the asymmetric information and the sorting effects are present. However, our conclusion is that the information effect is more relevant in the unmediated market, given the higher proportion of price decreases and the negative median and average resale prices. We do not read too much into the price changes after 3 months, as natural car depreciation is also likely playing a role, although the monthly depreciation rate implied by the resales is much steeper than what we observe in the general population of car transactions. Also, the fact that dealer sales rarely are resold immediately in the unmediated market (particularly for younger cars) suggests that adverse selection is not present in the dealer market.

## 5. Conclusion

■ Although there is a rich theory literature that connects product intermediation to product quality, the empirical literature on intermediation largely ignores this role, with two exceptions being Galenianos and Gavazza (2017) and Leslie and Sorensen (2013). We find evidence that used car dealers sell cars with higher quality, and we argue that these empirical regularities can be explained by theoretical models based on two prevailing views of intermediaries: dealers alleviate information asymmetry and dealers facilitate assortative matching in a frictional market. We also show that the data are more consistent with the theory of asymmetric information. We make a number of reduced-form assumptions to keep our model simple and focused; therefore, our model cannot be used to quantitatively decompose the different factors that lead to the dealer premium or to analyze the welfare consequence of car dealers. This is a natural direction for future work. Recent structural works by Salz (2017), Gavazza (2016), and Galenianos and Gavazza (2017) give us hope that this way forward is a possibility, although the addition of asymmetric information to these models would significantly complicate the analysis.

### References


Adams, C., Hosken, L., and Newberry, P. “Vettes and Lemons on eBay.” *Quantitative Marketing and Economics*, Vol. 9 (2011), pp. 109–127.

Adams, W., Einav, L., and Levin, J. “Liquidity Constraints and Imperfect Information in Subprime Lending.” *The American Economic Review*, Vol. 99 (2009), pp. 49–84.

Akerlof, G.A. “The Market for” Lemons”: Quality Uncertainty and the Market Mechanism.” *The Quarterly Journal of Economics*, Vol. 84 (1970), pp. 488–500.

Berndt, A. and Gupta, A. “Moral Hazard and Adverse Selection in the Originate-to-Distribute Model of Bank Credit.” *Journal of Monetary Economics*, Vol. 56 (2009), pp. 725–743.

Biglaiser, G. “Middlemen as Experts.” *The RAND Journal of Economics*, (1993), pp. 212–223.

Biglaiser, G. and Friedman, J.W. “Middlemen as Guarantors of Quality.” *International Journal of Industrial Organization*, Vol. 12 (1994), pp. 509–531.

——— and ———. “Adverse Selection with Competitive Inspection.” *Journal of Economics & Management Strategy*, Vol. 8 (1999), pp. 1–32.

Biglaiser, G. and Li, F. “Middlemen: the Good, the Bad, and the Ugly.” *The RAND Journal of Economics*, Vol. 49 (2018), pp. 3–22.

Bond, E.W. “A Direct Test of the ”Lemons” Model: The Market for Used Pickup Trucks.” *American Economic Review*, Vol. 72 (1982), pp. 836–840.

———. “Trade in Used Equipment with Heterogeneous Firms.” *Journal of Political Economy*, Vol. 91 (1983), pp. 688–705.

———. “Test of the Lemons Model: Reply.” *American Economic Review*, Vol. 74 (1984), pp. 801–804.

Cardon, J.H. and Hendel, I. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics*, Vol. 32 (2001), pp. 408–427.

Donna, J., Trindade, A., Pereira, P., and Pires, T. "Measuring the Welfare of Intermediation in Vertical Markets." In 2018 Meeting Papers, 984. Society for Economic Dynamics, 2018.

Einav, L., Finkelstein, A., and Schrimpf, P. "Optimal Mandates and the Welfare Cost of Asymmetric Information: Evidence from the UK Annuity Market." *Econometrica*, Vol. 78 (2010), pp. 1031–1092.

Engers, M., Hartman, M., and Stern, S. "Are Lemons Really Hot Potatoes?" *International Journal of Industrial Organization*, Vol. 27 (2009), pp. 250–263.

Galenianos, M. and Gavazza, A. "A Structural Model of the Retail Market for Illicit Drugs." *American Economic Review*, Vol. 107 (2017), pp. 858–96.

Gavazza, A. "An Empirical Equilibrium Model of a Decentralized Asset Market." *Econometrica*, Vol. 84 (2016),———

Gehrig, T. "Intermediation in Search Markets." *Journal of Economics & Management Strategy*, Vol. 2 (1993), pp. 97–120.

Genesove, D. "Adverse Selection in the Wholesale used Car Market." *Journal of Political Economy*, (1993), pp. 644–665.

Hayek, F.A. "The Use of Knowledge in Society." *The American Economic Review*, Vol. 35 (1945), pp. 519–530.

Hendel, I. and Lizzeri, A. "Adverse Selection in Durable Goods Markets." *American Economic Review*, Vol. 89 (1999), pp. 1097–1115.

Hendel, I., Nevo, A., and Ortalo-Magné, F. "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison. com." *The American Economic Review*, Vol. 99 (2009), pp. 1878–1898.

Hendren, N. "Private Information and Insurance Rejections." *Econometrica*, Vol. 81 (2013), pp. 1713–1762.

Hwang, I. "Dynamic Trading with Developing Adverse Selection." *Journal of Economic Theory*, Vol. 176 (2018), pp. 761–802.

Johri, A. and Leach, J. "Middlemen and the Allocation of Heterogeneous Goods." *International Economic Review*, Vol. 43 (2002), pp. 347–361.

Keys, B.J., Mukherjee, T., Seru, A., and Vig, V. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans." *The Quarterly journal of economics*, Vol. 125 (2010), pp. 307–362.

Lacko, J.M. "Product Quality and Information in the Used Car Market." Bureau of Economics Staff Report to the Federal Trade Commission, 1986.

Larsen, B. "The Efficiency of Real-World Bargaining: Evidence from Wholesale used-auto Auctions." Technical report. Cambridge, Mass.: National Bureau of Economic Research, 2014.

Leslie, P. and Sorensen, A. "Resale and Rent-Seeking: An Application to Ticket Markets." *Review of Economic Studies*, Vol. 81 (2013), pp. 266–300.

Lizzeri, A. "Information Revelation and Certification Intermediaries." *The RAND Journal of Economics*, Vol. 30 (1999), pp. 214–231.

Mian, A. and Sufi, A. "The Consequences of Mortgage Credit Expansion: Evidence from the US Mortgage Default Crisis." *The Quarterly Journal of Economics*, Vol. 124 (2009), pp. 1449–1496.

Murry, C. and Schneider, H. "The Economics of Retail Markets for New and Used Cars." In E. Basker, ed., *Handbook on the Economics of Retail and Distribution*. Cheltenham, UK: Edward Elgar, 2015.

Nosal, E., Wong, Y.Y., and Wright, R. "More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets." *Journal of Money, Credit and Banking*, Vol. 47 (2015), pp. 7–37.

———, ——— and ———. "Intermediation in markets for goods and markets for assets." *Journal of Economic Theory*, Vol. 183 (2019), pp. 876–906, Elsevier.

Peterson, J.R. and Schneider, H.S. "Adverse Selection in the used-car Market: Evidence From Purchase and Repair Patterns in the Consumer Expenditure Survey." *The RAND Journal of Economics*, Vol. 45 (2014), pp. 140–154.

Rhodes, A., Watanabe, M., and Zhou, J. "Multiproduct Intermediaries." Tech. rep., Yale University (2018).

Rubinstein, A. and Wolinsky, A. "Middlemen." *The Quarterly Journal of Economics*, Vol. 102 (1987), pp. 581–593.

Rust, J. and Hall, G. "Middlemen Versus Market Makers: A Theory of Competitive Exchange." *Journal of Political Economy*, Vol. 111 (2003), pp. 353–403.

Salz, T. "Intermediation and competition in search markets: An empirical case study." Available at SSRN 2961795, (2017).

Shevchenko, A. "Middlemen." *International Economic Review*, Vol. 45 (2004), pp. 1–24.

Spulber, D.F. "Market Making by Price-Setting Firm." *The Review of Economic Studies*, Vol. 63 (1996), pp. 559–580.

Wright, R. and Wong, Y.Y. "Buyers, Sellers, And Middlemen: Variations on Search-Theoretic Themes." *International Economic Review*, Vol. 55 (2014), pp. 375–397.

Yavaş, A. "Middlemen in Bilateral Search Markets." *Journal of Labor Economics*, Vol. 12 (1994), pp. 406–429.

———. "Search and Trading in Intermediated Markets." *Journal of Economics & Management Strategy*, Vol. 5 (1996), pp. 195–216.