

# Intermediaries and Product Quality in Used Car Markets\*

Gary Biglaiser      Fei Li      Charles Murry      Yiyi Zhou<sup>†</sup>

February 22, 2019

## Abstract

We examine used car dealers' roles as intermediaries. We present empirical evidence supporting that cars sold by dealers have higher quality: (1) dealer transaction prices are higher than private market prices and this *dealer premium* increases in the age of the car as a ratio and is hump-shaped in dollar value, and (2) used cars purchased from dealers are less likely to be resold immediately. We formalize a model to show that these empirical facts can be rationalized either when dealers serve to alleviate information asymmetry between sellers and buyers or when dealers facilitate assortative matching between heterogenous-quality cars and heterogeneous consumers. Lastly, based on predictions of the model, we use the data to distinguish these two theories and find evidence for both, but the preponderance of the evidence supports the asymmetric information theory.

**Keywords:** Adverse Selection, Sorting, Search Frictions, Car Dealer, Used Car, Intermediary, Middlemen

**JEL Classification Codes:** D82, D83, L15, L62

---

\*This paper supersedes the previous work titled "Middlemen as Information Intermediaries: Evidence from Used Car Markets." We 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, Katja Seim, Andrei Shleifer, Shouyong Shi, Senay Sokullu, Randy Wright, Andy Yates, Jidong Zhou, anonymous referees, participants of the 8th Annual Madison Meeting on Money, Banking and Asset Markets, 15th NYU IO day, 2017 FTC Micro Conference, 2017 SEAs, and seminar participants at Chinese University of Hong Kong, Hong Kong University of Science and Technology, Georgetown University, The Ohio State University, Shanghai University of Finance and Economics, and Stony Brook University for helpful comments.

<sup>†</sup>Gary Biglaiser: Economics Department, University of North Carolina at Chapel Hill, gbiglais@email.unc.edu. Fei Li: Economics Department, University of North Carolina at Chapel Hill, lifei@email.unc.edu. Charles Murry: Economics Department, Boston College, charles.murry@bc.edu. Yiyi Zhou: Department of Economics and College of Business, Stony Brook University, yiyi.zhou@stonybrook.edu.

# 1 Introduction

A vast majority of 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 the 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 (see [Yavaş \(1994\)](#), [Johri and Leach \(2002\)](#) and [Shevchenko \(2004\)](#)). Another popular justification of intermediary relies on frictions due to an informational asymmetry between agents. As argued by [Biglaiser \(1993\)](#) and [Lizzeri \(1999\)](#), intermediaries can serve as information intermediaries, or certifiers, in markets where there are motives for adverse selection or consumer sorting. The idea is that intermediaries have a 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 a number of distinct rationale for intermediaries, empirical research is limited and almost exclusively focuses 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 that car dealers provide high-quality cars for consumers, either motivated by information asymmetries or efficiency motives.

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

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

Based on these ingredients, we show that the empirical observations can be explained by two prevailing theories about intermediaries. First, we assume that the quality of the car is privately observed by the seller, but dealers are experts who can run a test to ascertain quality. The 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 direct private sales. The vintage of the car has an important effect on the dealer's price premium if age is correlated with the car becoming a lemon. On one hand, the dealer's information role increases as the car ages, but on the other hand, even high-quality cars depreciate naturally as they age. We show that this leads to a dealer premium pattern described above. In addition, the dealer's expertise of screening car quality generates a selection mechanism: cars purchased through dealers are more likely to be of high quality than direct transactions. By the classic adverse selection logic, buyers of lemons will resell them 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 market, cars with both high and low quality are sold privately 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 market: buyers with low valuations may purchase high-quality cars, giving them an incentive to resell their car. Therefore, the model also predicts that the resale rate is higher when the car is purchased from the 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. Specifically, the two theories make different predictions about the type of resold cars. In the asymmetric information theory, the expected quality of resale cars is lower than the expected quality of cars in the initial transaction because the resale is driven by buyers who want to get rid of a lemon. On the other hand, in the sorting theory, the expected quality of resale cars is higher because the resale takes place to improve the initial allocation. We show that resale rates are increasing in the age of a car at a faster rate for privately sold cars, and resale prices are more likely to be lower than the initial transaction price. Both of these facts support the information story being more important than the sorting story. However, the tests do not rule out either story, and we find evidence that the sorting mechanism may 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.

A number of 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 market is highly decentralized, and used cars are heterogenous, making the search and matching frictions in

the market non-trivial. Third, there are thousands of dealers per state, but many private-party transactions that are not intermediated, allowing us to compare the difference transaction patterns. Fourth, the used car market is large, with retail sales totally over 500 billion dollars annually in the United States.<sup>1</sup> In 2016, 38.5 million vehicles were sold in the second-hand market in the U.S., more than twice the number sold in the new car market.<sup>2</sup> Last, dealers are very active participants in the market. Nationally, about two-thirds of used car sales are made by dealers, and the other one-third occur between private parties. There are important differences between private sales and dealer sales. Private sales are much less regulated than dealer sales. Dealers are long-run players who sell many cars and care about their reputations, while private sellers are in the market very infrequently and have little reputation concerns. Furthermore, dealers are experts who may transact the same type of cars many times, and who employ mechanics on site.

## 1.1 Contribution and Related Literature

We present empirical evidence of the quality provision role of intermediaries that is 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 but not quality provision. Recent examples include Hendel, Nevo, and Ortalo-Magné (2009), Gavazza (2016), Salz (2017), and Donna, Trindade, Pereira, Pires, 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, both dealers and individuals facilitate trade in the used car market, and we examine both an information asymmetry story and an assortative matching story.<sup>3</sup> Our work is closely related to many studies on adverse selection, sorting and market segmentation, and intermediaries.

**Intermediaries.** The theoretical foundations of this paper 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),

---

<sup>1</sup>This number, constructed from Edmunds' and Manheim's yearly reports, represents revenues from franchised and independent dealers, so it is a conservative reflection of the size of the industry. We found conflicting reports about the total revenues of the private party sector.

<sup>2</sup>Our general understanding of the industry is 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 [https://dealers.edmunds.com/static/assets/articles/2017\\_Feb\\_Used\\_Market\\_Report.pdf](https://dealers.edmunds.com/static/assets/articles/2017_Feb_Used_Market_Report.pdf) and <https://publish.manheim.com/content/dam/consulting/2017-Manheim-Used-Car-Market-Report.pdf>

<sup>3</sup>A similar mechanism appears in Galenianos, Pacula, and Persico (2012), although in contrast to the used cars market, drug markets are characterized by repeated searches by consumers, so the exact mechanisms are different from ours. Another exception is Leslie and Sorensen (2013), who examine the allocative benefits of event ticket resellers, although information asymmetry is not present.

intermediaries emerge to identify lemons. Second, there is a large literature discussing the function of intermediaries to save search costs of agents in the 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. Lastly, there is a literature emphasizing the role of intermediaries to facilitate 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.<sup>4</sup> 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.<sup>5</sup> [Gavazza \(2016\)](#) shows that dealers reduce trading frictions through costly intermediation, but also impose an externality by crowding out the number of direct transactions. In other industries, [Hendel, Nevo, and Ortalo-Magné \(2009\)](#) compare house sales on a For-Sale-By-Owner (FSBO) on-line platform to the Multiple Listing Service (MLS), 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 intermediaries.

**Testing for Adverse Selection.** 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. Recently, inspired by the test derived by [Hendel and Lizzeri \(1999\)](#), [Peterson and Schneider \(2014\)](#) considered a car as an assemblage of parts, some with asymmetric information, and others without, and found evidence of adverse selection and consumer sorting. We contribute to the literature by comparing the transaction price, conditional on age, of dealers with those in direct sales in the entire market. Also, rather than testing for the presence of asymmetric information by examining

---

<sup>4</sup>Relatedly, [Kim \(2012\)](#) and [Guerrieri and Shimer \(2014\)](#) show that decentralized markets under adverse selection and search frictions can be endogenously segmented in a way that improves social welfare. Endogenous segmentation is driven by low-quality sellers’ incentive to attract more buyers by separating from high-quality sellers.

<sup>5</sup>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 private sellers, although this is not their primary focus.

sellers' adverse selection, we focus on the selection made through dealers.<sup>6</sup>

The idea of using turnover rates to proxy quality has been widely applied 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 attributed as evidence of adverse selection. See, e.g., [Berndt and Gupta \(2009\)](#), [Mian and Sufi \(2009\)](#) and [Keys, Mukherjee, Seru, and Vig \(2010\)](#). We focus on resales taking place within a short time period after the initial transactions to tease out the reallocation resulting from depreciation, which plays a central role in [Bond \(1983\)](#), [Hendel and Lizzeri \(1999\)](#) and [Peterson and Schneider \(2014\)](#).

The rest of the paper 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 private 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 two theories. Section 5 concludes.

## 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 market.

**On Quality of Cars.** Cars are highly differentiated products and very complicated machines. 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 to 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, things we do not consider quality are features like car age (directly), mileage, or make/model/trim.

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 a seller. For example, a car may have visible exterior/interior damage or the owner may have receipts from maintenance, oil

---

<sup>6</sup>There is also a substantial literature on asymmetric information in other industries, particularly health and insurance markets, including [Cardon and Hendel \(2001\)](#), [Einav, Finkelstein, and Schrimpf \(2010\)](#), and [Hendren \(2013\)](#). The institutions in these markets are somewhat different, and "buyers" (insurance providers) have focused on pricing mechanisms based on observable information, for example, credit scores and demographic information. Our impression is that the role of intermediaries that screen asymmetric information is very limited in these markets.

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.<sup>7</sup> Our first goal is to present evidence that used car dealers offer higher-quality cars than private sellers.

**Empirical Strategy.** Our challenge is that, by definition, quality is unobservable to the econometrician 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.<sup>8</sup> First, according to the efficient market hypothesis, the market price aggregates dispersed information and reflects the expected value of traded products. In the used car market, if dealers sell higher-quality cars than the private market, one should expect that dealers enjoy a positive price premium relative to the market for cars with observably identical characteristics. However, a positive price premium does not necessarily imply that dealer cars have superior quality because, in addition to selling the product *per se*, the dealer provides a sequence of pre-transaction 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, but they do affect the buyers' payoff and therefore their shopping decision and willingness to pay.<sup>9</sup>

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. (1) 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. (2) 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. Hence, 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 pre-transaction service to generate the age pattern on the price premium.

Second, we examine the *post-transaction* resale rate of cars. If dealer sell higher-quality cars, cars should be quickly resold more often than if they are from the private market. The reason is

---

<sup>7</sup>Peterson and Schneider (2014) elaborate on this distinction between observed and unobserved quality using repair services for particular parts of the car.

<sup>8</sup>We formalize the following theoretical arguments in Section 3.

<sup>9</sup>Many other factors may also contribute to the price premium, e.g., (i) underreported price in private transaction for tax avoidance, (ii) bargaining power difference between dealers and private sellers, etc. What is important for our empirical strategy is that these other factors do not correlate with the age of a car.

twofold. (1) A buyer may be uncertain about the quality of the car. When she realizes that her purchase is a lemon, she will be more likely to resell it. (2) The initial allocation in the private market may be less efficient than through the dealer due to buyer heterogeneity in willingness to pay for quality and differential search and matching frictions. On the one hand, the dealer chiefly trades high-quality cars at higher prices, which mainly attracts buyers with high valuations, leading to a more efficient allocation. On the other hand, in the private market, transactions are less organized, information is less aggregated, and car quality is more dispersed, so an inefficient allocation is more likely to occur. In this case, reallocation takes place to “correct” the initial allocation from the private market. A buyer who purchased from the private market will be more likely to meet another agent who has a higher valuation for the car, leading to higher resale rates for cars traded in the private market.

**Empirical Results.** First, using used car registration data from the Virginia Department of Motor Vehicles, we show that dealer sales have higher transaction prices than transactions between private parties. Furthermore, we document that the dealer price premium is increasing with the car age in percentage terms, and is hump-shaped in car age in dollar terms. Using Pennsylvania used car registration data, we show that used cars purchased from dealers are less likely to be re-sold in a short time frame. We conclude that this empirical evidence strongly supports the hypothesis that there exists a quality difference between cars sold by dealers and those in the private market.

## 2.1 Price Premium of Dealers

### 2.1.1 Used Car Registration Data from Virginia

We analyze the universe of used car registrations from 2007 to 2014 in Virginia and document the difference between transaction prices of dealers and private sales, and the car age patterns of this difference. The dataset was obtained from Virginia Department of Motor Vehicles (VA-DMV), and it includes all used car transactions registered in Virginia from January 1, 2007, to December 31, 2014. For each registration, we know the transaction date, price,<sup>10</sup> the first 12 digits of the Vehicle Information Number (VIN) which is a unique number assigned to a vehicle that

---

<sup>10</sup>The price of the 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. For example, Toyota and Honda’s CPO program (from their new dealers) includes a one-year warranty. Other warranties that a consumer can purchase are not included in this price. We collected data from cars.com in 2015 and can report that about one-third of cars 4 years old and younger have a CPO designation, while older cars are rarely sold as CPO. According to *Edmunds*, about 7 percent of all used car transaction are CPO cars. We conduct a robustness check in our analysis by excluding young cars and excluding used cars from new car dealers (CPO programs are offered through manufacturers, so they are available only at dealers who have new car franchises).



contains information to describe and identify the vehicle,<sup>11</sup> 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. We merge the dealer identification numbers with a separate dataset provided by the DMV that includes identification numbers matched to dealer names and addresses. Buyers are also marked as “private buyers” or with dealer identification numbers. 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 edmunds.com, 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 DMV data with a list that matches zip codes to counties.

We make a number of sample selection decisions for the raw data to focus on our research questions. First, we drop 387,926 transactions when dealers are buyers.<sup>12</sup> Second, we discard transactions with negative odometer readings and cars 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. In the end, our sample includes 5,469,241 transactions. Among them, 3,286,326 transactions (60 percent) were made by car dealers and the remaining 2,566,349 (40 percent) were made by private sellers.

Table 1 presents summary statistics from our sample, including the transaction price, car age, and odometer mileage for the two segments. Overall, cars sold by dealers were substantially newer and more expensive than those sold by private sellers. Specifically, an average dealer car was around 6 years old and sold at a price of \$13,032, whereas an average non-dealer car was 11 years old and sold at a price of \$3,960. However, the standard deviations of car age and transaction price are large, indicating that there was substantial heterogeneity across transactions.

---

<sup>11</sup>The VIN standard, created by the National Highway Traffic Safety Administration (NHTSA) and enforced starting with the model year 1981, was required of all vehicles manufactured for use in the U.S. The NHTSA requires the VIN to be 17 digits long. The first three digits are reserved for the World Manufacturer Identification number and identify the manufacturer and country of origin of the vehicle. The fourth to eighth digits capture descriptive elements of the vehicle, including engine, body type, drive type, doors, restraint system and Gross Vehicle Weight (GVW) range. The ninth digit is a check digit that can be used to verify the validity of an encountered VIN using a calculation. The tenth digit identifies the model year of the vehicle and the eleventh digit identifies the specific plant and plant location where the vehicle was manufactured. The twelve to seventeen digits are serial numbers.

<sup>12</sup>Among them, 171,634 transactions were between dealers and 216,292 transactions were made from individual sellers to dealers. Dealers do not necessarily need to re-register a car, so these observations do not represent the universe of dealer purchases.

Table 1: Summary of Virginia DMV Data

	Mean	SD	Q25	Q50	Q75
<b>Private Sales</b>					
Price	3,960	5,144	1,000	2,000	4,500
Mileage	134,376	67,290	92,183	132,315	171,300
Car Age	11.14	4.38	8	11	14
<b>Dealer Sales</b>					
Price	13,032	8,518	6,349	12,000	17,779
Mileage	77,402	53,325	36,449	66,675	107,811
Car Age	5.99	4.05	3	5	9
<b>Dealer Sales: 60.09%</b>					
<b>Total Transactions: 5,469,241</b>					

Note: The 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.

Figure 1 presents the total transactions of the two segments across different car vintages. First, the total number of dealer transactions falls in car age after peaking at three-year-old cars, which is the common lease length for leasing cars. Second, the total number of transactions sold by private sellers increases in car age until age twelve and then falls in car age. We also graph the share of dealer sales by vintage, which is strictly decreasing with car age.<sup>13</sup> We also merge our transaction data with the Census data to get the local demographics at the buyer’s zip code. Figure A.6 shows that there is a positive correlation between the dealer share and the median household income at the buyer’s zip code.

Next, we describe the data in terms of the most popular brands. We list descriptive statistics of the ten most popular brands in the data in Table 2. Most of the top ten brands are common U.S. and Japanese brands, with Ford and Chevrolet combining for 27% of the transactions and Honda and Toyota combining for 20% of transactions. The only luxury brand in the top ten is BMW, at number ten with 3% of the transactions in the data. The aggregate patterns in the data hold across all the brands: the dealer share of transactions is over half, average dealer prices are much higher than direct transactions prices, and dealer sales typically involve younger cars than private transactions.

Lastly, we summarize the prices of used car transactions from dealers and direct sales for every car age. We plot the average transaction price by car age in the left panel of Figure 2. The two

<sup>13</sup>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.

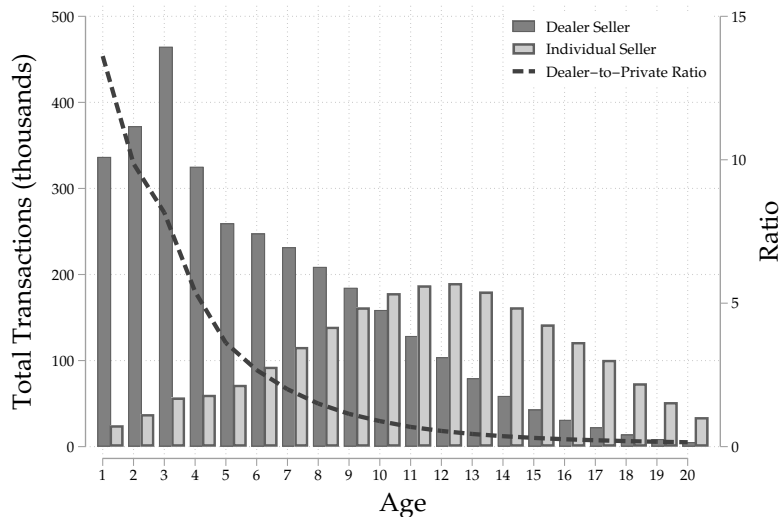


Figure 1: Dealer and Private Sales

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

Table 2: Summary of Transactions, by Brand

Brand	Market Share	Transactions			Mean Price		Mean Age	
		Total	Dealer	Dealer Share	Dealer	Direct	Dealer	Direct
Ford	15%	794,677	448,338	56%	11,837	3,470	6.26	11.14
Chevrolet	12%	629,347	388,996	62%	12,281	3,943	5.90	10.65
Honda	10%	541,635	269,920	50%	12,116	3,426	6.35	12.28
Toyota	10%	534,206	307,176	58%	13,930	4,479	5.62	11.45
Nissan	7%	357,329	226,071	63%	12,785	3,500	5.42	11.37
Dodge	5%	296,554	194,970	66%	11,829	3,908	5.56	9.88
Jeep	3%	187,788	119,694	64%	13,114	4,160	6.04	11.24
Volkswagen	3%	141,306	86,043	61%	11,413	4,174	5.87	9.99
Chrysler	3%	138,432	98,788	71%	11,275	3,684	5.29	9.56
BMW	3%	137,132	93,275	68%	21,209	8,540	5.81	10.66

Note: The 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.

downward-sloping lines are the transaction prices for dealer sales and private sales. The upward-sloping line (associated with the right axis) is the ratio of these two prices. Dealer prices are higher than direct prices at every age. The *difference* in the average prices increases at first, and then decreases, so that very old cars have similar average prices. The *ratio* of prices 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 from dealers and direct sales may differ across vintages due to compositional effects, and the following empirical analysis will control for these compositional changes by using within trim variations in prices. In the remainder of the empirical analysis, we examine how prices are correlated with age, but it could also be the case that mileage is the primary consideration when thinking about the asymmetric information 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. We display the average transaction prices by mileage in the right panel of Figure 2.

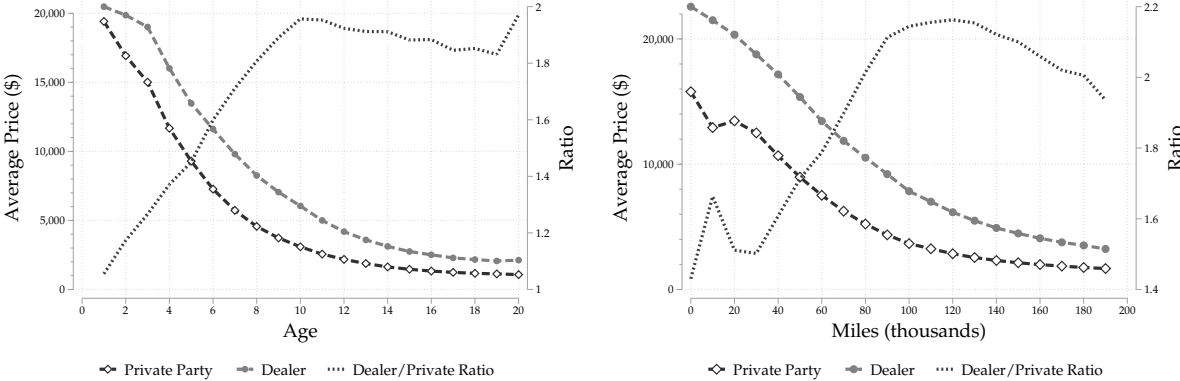


Figure 2: Transaction Prices

Note: Mean transaction prices by car age (left panel) and car mileage (right panel). An observation is a single used-car transaction in Virginia from 2007 to 2014. The sample is described in the text.

**2.1.2 Dealer Price Premium and Age Effect**

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 private market, conditional on observed car characteristics (observed by the econometrician) 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 private 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 seller, indicators for different car ages, and age indicators interacted with the indicator of dealer seller. 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 seller and car age indicators capture to what extent the dealer price premium co-varies with car age. Essentially, we compare prices of two observationally equivalent cars (same model, same model year, same trim, same odometer mileage, and vintage), with one being sold at dealer and other one being sold by a private seller, and we examine how this price difference varies in car age.

In specification (1), we include all used car transactions in our sample described above except for those extremely unpopular products with fewer than 100 transactions over the eight years (from 2007 to 2014) which account for less than 2% of the sample. We are left with 5,325,273 transactions, representing 35,248 unique model-model year-trims. 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 private sales and dealer sales from used-car-only dealers who do not have new car business lines. 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. Lastly, to reduce the potential impacts of leasing cars, rental cars, CPOs, and substitution from new cars, in specification (4) we only include transactions that include cars that are at least four years old.

The estimation results are reported in Table 3 and Figure 3. The estimates are extremely precise, with every coefficient we report being statistically significant at least at the 0.001 level, using robust standard errors. As expected, the coefficient for the log of mileage is negative.<sup>14</sup> The coefficients and associated standard errors for car age indicators are reported graphically in Figure 3a. 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 four years old rather than one year old in other specifications. The coefficients and associated confidence intervals for the age-dealer interactions are graphically reported in Figure 3b. The interaction coefficients are precisely estimated, and increase monotonically until age ten and thereafter level off and fall

---

<sup>14</sup>In an alternative specification we included dummies for mileage bins, as in Peterson and Schneider (2014), and our results are nearly identical.

slightly.

Table 3: Dealer Premium Regressions

	(1)	(2)	(3)	(4)
log(Mileage)	-0.286	-0.326	-0.311	-0.375
Constant	12.553	12.904	12.736	13.098
Age Effects	..... See Figure 3a .....			
Age-Dealer Interactions	..... See Figure 3b .....			
$R^2$	0.750	0.471	0.547	0.460
Num. Observations	5,325,273	3,600,473	1,156,736	4,091,603

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 seller indicator. All point estimates are statistically significant at least at the 0.001 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 four years old.

Based on the estimates, we compute the *predicted dealer premium* as a difference in dollars across different car ages and display the results in Figure 4a.<sup>15</sup> For all specifications, the age profile of the average dealer premium is hump-shaped and reaches its peak at age six, at a value of between \$3,500 and \$4,000, depending on the specification.<sup>16</sup> This is a large premium given that the average price of a six-year-old dealer car is roughly \$12,000 (see Figure 2). After age six, the price premium declines monotonically until age twenty (less than \$1,000). Moreover, we compute the predicted dealer premium ratio by car age and display the results in Figure 4b. The price ratio of dealer sales over private sales is increasing in car age until age ten, with a value of approximately 2 at that age, and then flattens and decreases slightly after age ten. It is not surprising that our estimates are noisier for older cars, since dealer sales dropped substantially for old cars; see Figure 1.

To summarize, our data suggest 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.*

**Robustness** To control for those unobserved local factors affecting used car prices, we estimate the four specifications by including seller county effects, and present the predicted dealer price premiums across different car ages in Appendix A.1. The results are very similar to those shown

<sup>15</sup>Note that since our dependent variable is log price, this involves a non-linear transformation of the estimates. The standard errors are adjusted accordingly.

<sup>16</sup>We repeat the analysis estimating the regression with price levels as the dependent variables, as opposed to logs. The results are in Appendix A.3.

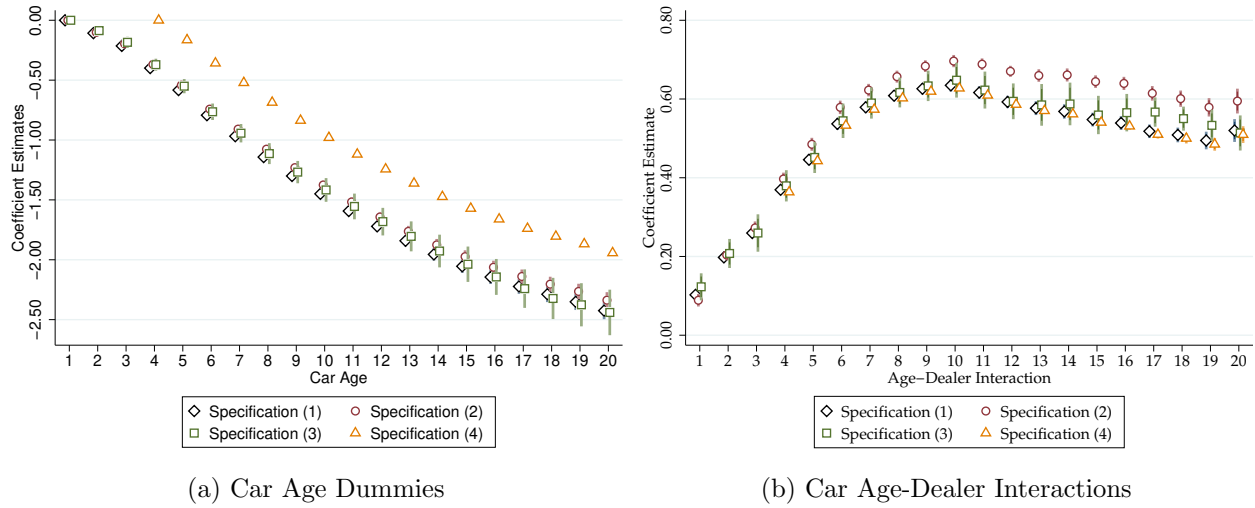


Figure 3: Coefficient Estimates

Note: Point estimates with 99% confidence intervals. Different specifications refer to the different columns in Table 3.

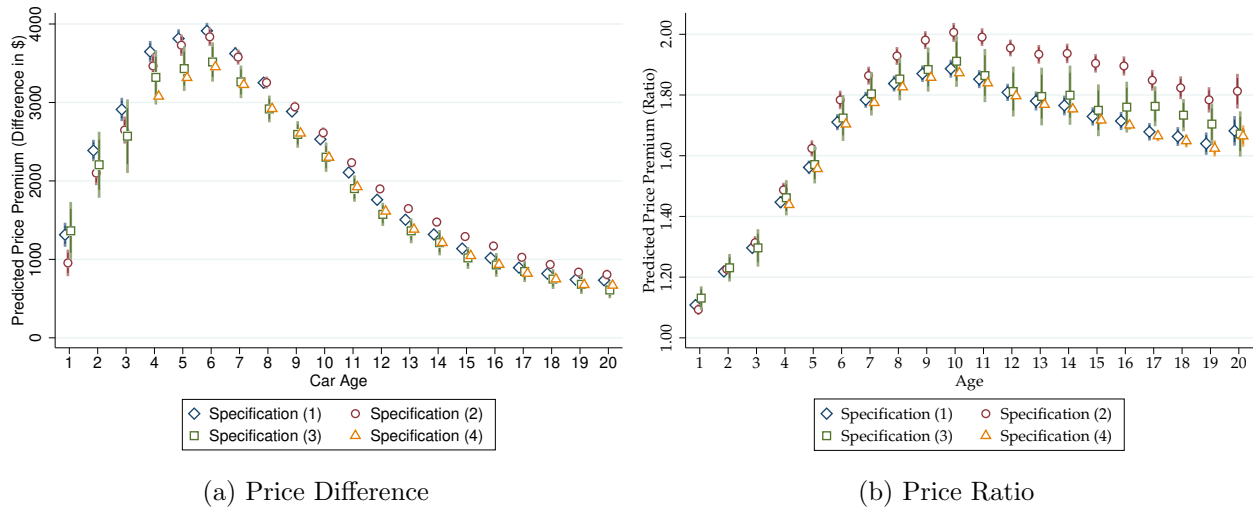


Figure 4: Predicted Dealer Premium

Note: Point estimates with 99% confidence intervals. Different specifications refer to the different columns in Table 3.

in Figures 4a and 4b.<sup>17</sup> We also merge our data with information from *Consumer Reports* which provides model and model-year level ratings of the reliability of many cars in our sample. We examine the dealer premium by different levels of car reliability. In other words, we can rank the age shape of dealer premium by how reliable is the car. Details of this analysis can be found in Appendix A.3. Again, dealer price premium in percentage terms is increasing in car age, regardless of the reliability rating of the car models. We also re-estimate the hedonic price regression by replacing the log price with the price level as the dependent variable, and present the results in Appendix A.4. The results are similar to Figure 4.

**Matching Estimator.** We estimate the dealer price premium using a matching estimator. To implement the matching estimator, we exactly match dealer and private cars on the following variables: make, model, trim, model year, mileage, and seller county, where we create coarse bins for mileages (we use bins of 30k miles as in Peterson and Schneider, 2014). In general, the results look very similar to those of our main fixed effects regression analysis. We discuss the specifics in Appendix A.2 and we present the results in Figure A.2.

**Price Dispersion.** Lastly, we document that the dealer premium is not just an average effect, but the entire distribution of dealer prices first-order stochastically dominates the distribution of private market prices. To do this, we run a hedonic price regression with model-trim-model year fixed effects, similar to the regression from Table 3, but *without the dealer dummy*. In Figure 5, we plot the empirical cumulative distribution function of the standardized residuals from this regression for dealer and private market cars, separately. The dispersion in residualized prices is less for old cars no matter what the source, and the distributions for old cars look more similar than the two distributions for young cars. In both cases, we can easily reject the null that the two distributions are the same using a Kolmogorov-Smirnov test.

## 2.2 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 Virginia DMV 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-DOT). It covers all used car transactions registered from January 1, 2014, to July 31, 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, compared to

---

<sup>17</sup>Summary statistics of this sample are in Table 7, in Appendix A.1.



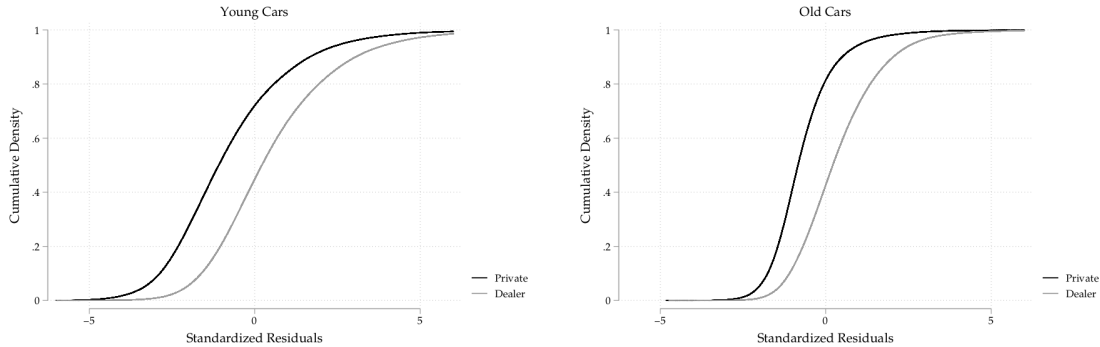


Figure 5: Cumulative Distributions of Residualized Prices

Note: Young = 3-6 years old. Old = 7-10 years old.

the Virginia data, the time panel is substantially shorter, so the comparative advantage of the data is testing our resale hypothesis.

### 2.2.1 Used Car Registration Data from Pennsylvania

The Pennsylvania data include 2,339,102 used car transactions with cars no more than 20 years old. Among them, 54% of cars were sold by dealers and the remaining 46% were sold by private sellers. 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 1,430,307 unique cars transacted during this period, with 761,867 cars (53%) being sold by dealers.

We define a resale as a VIN that appears multiple times in our Pennsylvania transactions dataset. Among all 1,430,307 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 end up with 90,911 resales that occurred between January 2014 and before 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.<sup>18</sup>

### 2.2.2 Resale Rates: Dealer Sales versus Private Sales

Table 4 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

<sup>18</sup>We also conduct our analysis with the original 153,892 resale transactions and find very similar results.

buying from a dealer and buying from a private seller. Regardless of the post-transaction time windows, the resale rates of dealer cars are substantially lower than those of cars sold by private sellers. For example, 0.52 percent of dealer cars were resold within one quarter after transaction, in contrast to 2.13% of cars sold directly by private sellers.

Table 4: Summary Statistics, Resales after Purchase

	Dealer Sales		Direct Sales	
No. of Initial Sales	719,606	(53%)	647,720	(47%)
Resale within one quarter	3,729	(0.52%)	13,775	(2.13%)
Resale within two quarters	7,308	(1.02%)	22,862	(3.53%)
Resale within three quarters	11,269	(1.57%)	31,236	(4.82%)
Resale within four quarters	15,707	(2.18%)	39,896	(6.16%)

Note: Percentage of used car sales that were resold after one, two, three, and four quarters.  
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 for cars’ observable characteristics, analogous to our empirical strategy of the price regression:

$$y_i = 1\{\mu_i + \beta_d d_i + \mathbf{x}_i \boldsymbol{\beta}_x + \epsilon_i > 0\} \quad (1)$$

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,  $\mathbf{x}_i$  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 5 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

Table 5: Immediate Resale after Purchase: Logit with Product Fixed Effects

	Resale Time Window			
	One Quarter	Two Quarters	Three Quarters	Four Quarters
Bought from Dealer	-0.761 (0.022)	-0.615 (0.016)	-0.532 (0.013)	-0.478 (0.009)
Log Mileage	0.238 (0.022)	0.278 (0.017)	0.283 (0.014)	0.292 (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. Sample selection is described in text. Source: Pennsylvania Department of Transportation.

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.

### 2.2.3 Sample with Dealer Inventory

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, 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 sub-prime 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 private sales, so it should be correlated with  $d_i$ . High levels of inventory may also put downward pressure on prices in local markets because (1) tighter competition across dealers and (2) the opportunity costs associated with inventory capacity for a particular dealer lot. On the other hand, it is unlikely that initial inventories are an important determinant of whether a buyer resells many weeks later.<sup>19</sup>

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.<sup>20</sup>

Table 6 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

---

<sup>19</sup>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.

<sup>20</sup>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.

bottom panel of Table 6, 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.

Table 6: Summary Statistics: Dealer Inventories

	Mean	SD	Q25	Median	Q75
Dealer-Week Inventories	55.15	55.63	19	41	75
– <i>Convertible</i>	2.00	1.59	1	1	2
– <i>Coupe</i>	3.01	2.56	1	2	4
– <i>Hatchback</i>	4.41	4.25	2	3	6
– <i>Minivan</i>	4.06	5.40	1	3	5
– <i>SUV</i>	21.84	23.52	7	16	31
– <i>Sedan</i>	24.08	25.53	8	17	32
– <i>Wagon</i>	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.

**Threats to identification** Our instrument relies on the assumption that current inventories do not affect the buyer’s decision to resell quickly – up to six 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 are wholesale auctions. Many dealers, particularly dealers with new-car franchises, rely on trade-ins as well. Trade-ins to the latter are often either re-sold at auctions to other dealers traded to commonly-owned dealers.<sup>21</sup> In Figure 6 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 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 if there is 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 to the aggregate new-car shock. However, the patterns in Figure 6 do not seem consistent with this aggregate shock story,

<sup>21</sup>See Larsen (2014) and Murry and Schneider (2015) for details.

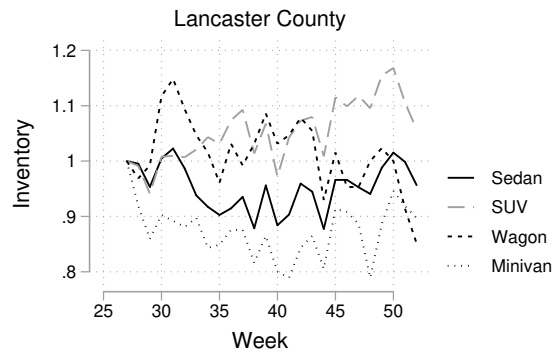
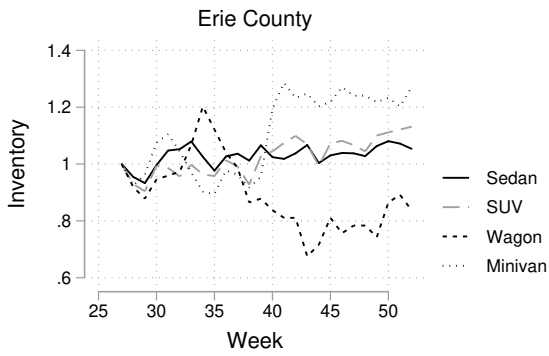
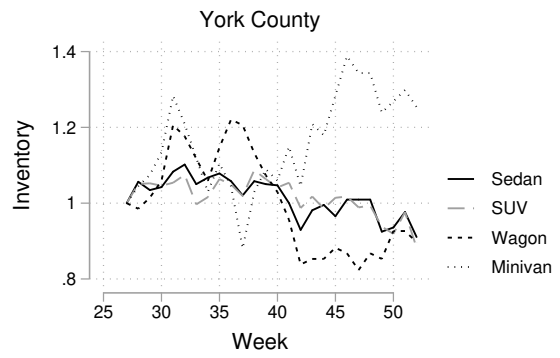
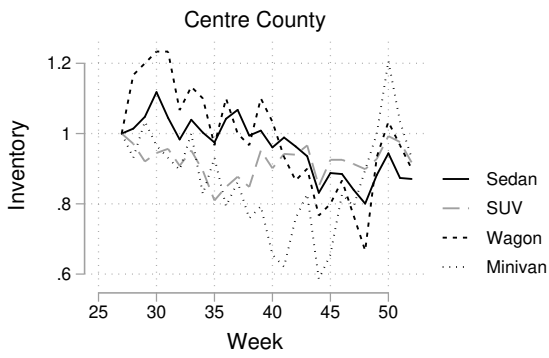


Figure 6: County Inventories by Style

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 auto auctions that are due to institutional reasons as opposed to consumer preferences.

**CarMax** We also instrument for dealer sales using the relative locations of dealers to a CarMax location. CarMax is a large national chain and typically has one of the largest inventories in a given local area and a very large virtual inventory because they 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.

#### 2.2.4 Results of Control Function Approach

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 the column (I) of Table 7. The estimate of the coefficient before the excluded variable is positive and significant at 10 percent 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 8. The first two columns are the results for the Logit model with model-year-trim-car age fixed effects, and the last two columns are the results for the control function approach. Again, cars bought from dealers are less likely to be resold shortly after purchase, with the effect being stronger for the first quarter than two quarters. The estimates of the dealer seller coefficient using the control 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, and report the first stage results in the column (II) of Table 7. The estimate of the coefficient before the excluded variable is negative and significant at the 5 percent 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. Table 9 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.

Table 7: First-Stage Results

	(I)	(II)
Log of Inventory of Nearby Dealers	0.012 (0.007)	- -
Log of Distance to the Nearest CarMax	- -	-0.022 (0.007)
Log Mileage	-0.414 (0.023)	-0.385 (0.021)

Note: The dependent variable is an indicator that indicates whether the car was bought from a dealer. All specifications include model-model year-trim-car age fixed effects, weekly dummies, and county dummies. In column (I), the excluded variable is the log of the inventory of dealers that locate in the same zip code as the buyer, of cars that have the same body style as the transacted car, during the week when the transaction occurred. In column (II), the excluded variable is the log of distance between the buyer and the nearest CarMax store. Standard errors in parentheses. The sample includes 72,538 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.

Table 8: Immediate Resale after Purchase: Logit with Control Function

	Fixed Effects Logit		Control Function	
	Resale Window		Resale Window	
	One Quarter	Two Quarters	One Quarter	Two Quarters
Bought from Dealer	-0.908 (0.096)	-0.765 (0.070)	-0.924 (0.103)	-0.781 (0.075)
Log Mileage	0.380 (0.108)	0.459 (0.082)	0.156 (0.337)	0.348 (0.216)

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. In the control function panel, we use dealer inventory as the excluded variable for whether a car was bought from a dealer. Standard errors in parentheses. The sample includes 72,538 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.

Table 9: Robustness Check: An Alternative Instrument Variable

	Resale Time Window	
	One Quarter	Two Quarters
Bought from Dealer	-0.921 (0.095)	-0.767 (0.070)
Log Mileage	0.666 (0.270)	0.906 (0.209)

Note: The dependent variable is an indicator for post-purchase resale within the specified time window. All specifications include model-year-trim fixed effects, weekly dummies, and county dummies. We use a Logit model with product fixed effects to model the first stage but use the log of the distance to the nearest CarMax as the excluded variable for whether a car was bought from a dealer. Standard errors in parentheses. The sample includes 72,538 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.

**Fact 2.** *Cars purchased from dealers are less likely to be immediately resold than privately purchased cars.*

### 2.3 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 private 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. On the other hand, although dealers' pre-transaction 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 privately also indicates quality differences between dealer and privately sold cars. Intuitively, the dealers' pre-transaction service should have very limited impact on buyers' post-transaction decisions if the quality distribution of cars sold in the two markets (dealer and private) are identical. In the next section, 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 are available from private 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. We will focus on two selection mechanisms based on different sources of market frictions, which are the two most prevailing roles of intermediaries in the literature. The model is deliberately simple but captures the most salient features of the used car market. In Section 3.1, we describe the basic ingredients of these models. In Section 3.2, we introduce *asymmetric information* into the model: a car’s quality is privately known by the seller and the dealer. To highlight the effect of information asymmetry, we assume buyers are *homogenous* and they do not know the true quality of a particular car. In this setting, the dealer serves as information intermediary, and obtains profits by selecting and selling high-quality cars. We derive empirical implications for how the price premium changes as the car ages and on the difference between resale rates of cars sold through dealers and private transactions. In Section 3.3, we examine the model with *complete 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. We show that many of the empirical implications in this latter setting are similar to the model with asymmetric information. In the appendix, we discuss the sensitivity and validity of our assumptions at length.

#### 3.1 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 sub-market in isolation.

**Dynamics of Car Quality.** 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.

**Sellers.** A seller remains passive until he receives a liquidity shock which arrives at a rate  $\mu$ . A seller must sell his car upon the arrival of the liquidity shock.<sup>22</sup> The car’s vintage,  $t$ , is *publicly observed*. 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:

$$\dot{q}_t = -\lambda_t q_t < 0, \forall t, \tag{2}$$

---

<sup>22</sup>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.

with the initial condition  $q_0 = 1$ .

For simplicity, 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 buyers directly if either he 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. What matters is that it ensures that some high-quality cars will be traded in the market. A seller's payoff equals the transaction price if he sells the car and zero, otherwise.

**Buyers.** 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  $\psi_t \in (0, 1)$ . 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  $\lim_{t \rightarrow \infty} U_t^H = 0$ , 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.

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 market.

**Dealer.** 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 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.

**Timing.** 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 are equal and normalize it to one. Thus, we examine each cohort of cars in isolation.

The order of moves of cohort  $t$  game is given as follows:

1. 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 private market.
2. The dealer makes a take-it-or-leave-it selling offer to each buyer. Each buyer decides between accepting the offer and rejecting it and going to the private market.
3. In the 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.

### 3.2 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. 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.<sup>23</sup> 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  $q_t$ .<sup>24</sup> When a buyer and a seller meet in the market, the buyer observes neither the quality of the car nor whether the seller has visited the dealer.

We analyze players' incentives via backward induction. We begin with the transaction in the private market. Because  $\theta_t$  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 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.<sup>25</sup> The seller rationally anticipates his payoff is  $b_t$  if he goes to the 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. Because  $k > U_0^H$  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

---

<sup>23</sup>Our result is robust to the extension where the dealer observes an informative signal about the quality.

<sup>24</sup>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  $t \rightarrow \infty$ ,  $q_t \rightarrow 0$ . See [Hwang \(2018\)](#) for a more detailed discussion of developing asymmetric information.

<sup>25</sup>Buyer beliefs off-equilibrium path that assume any different offer comes from a low-quality seller are sufficient for this.

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 private market only if the seller fails to find the dealer; and thus in the equilibrium,

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

The numerator is the measure of high-quality cars directly sold in the 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; while 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 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 market, a buyer accepts the seller's offer if and only if the price is not higher than  $b_t$  satisfying (3) in the 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 private market. Accordingly, the buyers will lower their belief of the quality of cars on the private 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 private 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 private market reflects two 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, since 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 market, which further enlarges the quality difference between the dealer's supply and the supply on the private market.

**Price Premium Dynamics.** Fixing the car's vintage and other observable characteristics, we call the difference in the transaction price at the dealership and the market the dealer *price premium*. The dealer's price premium varies as the car ages. Although both the dealer price,  $U_t^H$ , and the 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 a direct 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 the previous section.

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

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

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

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

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 market, lowering the 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  $t \rightarrow \infty$ ,  $U_t^H$  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 direct sales in percentage terms:

$$\frac{p_t}{b_t} = \frac{1/q_t - \alpha}{1 - \alpha}. \quad (6)$$

By taking the ratio between the dealer transaction price and direct 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  $t$ .<sup>26</sup>

Formally, we summarize our first empirical implication 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 direct sales in percentage terms formulated in (6) is greater than one for all car ages and is increasing in the car’s age.*

The implication is consistent with Fact 1.

An alternative way to understand the dynamics of premium is to compare the “declining rate” between direct sale price and dealer price as in Hendel and Lizzeri (1999). The direct sale price declines at a rate

$$\frac{\dot{b}_t}{b_t} = \frac{\dot{U}_t^H}{U_t^H} + \underbrace{\frac{\dot{q}_t}{q_t} + \frac{\alpha \dot{q}_t}{1 - \alpha q_t}}_{(-)},$$

which is faster than the price declining rate of the dealer price  $\dot{p}_t/p_t = \dot{U}_t^H/U_t^H$ . The falling dealer price is driven by the depreciation effect only, while the reduction in the market price also reflects the fact that older cars are more likely to be lemons.

**Resales.** 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, 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 private seller. For simplicity, we do not explicitly modeling the demand and the transaction process of the resale market, and we assume the resale market is competitive and price

---

<sup>26</sup>The fact that the depreciation effect drops out in the ratio is the consequence of the separability of the depreciation effect and the asymmetric information effect, which depends on neither the normalization  $U_t^L = 0$  nor the binary type assumption. 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  $\lim_{t \rightarrow \infty} U_t = 0$ . The term  $U_t$  captures the separable depreciation effect. Then the  $U_t$  still drops out in the price premium in percentage terms.

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, while 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

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

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.

### 3.3 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. That is, we assume that the car's quality  $\theta_t$  is observed by all players but not the econometricians, and  $\phi < 1$ , the high-valuation consumers value a high quality car more than the low-valuation consumers. 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 is able to selectively attract high-valuation buyers and sellers with high-quality cars through its pricing since it improves the average matching efficiency. The dealer's price premium reflects not only high quality of cars but the additional matching surplus being created. To avoid rationing and simplify the analysis, we assume that

$$\psi_t = \alpha q_t.$$

which implies that the maximum quantity of high-quality cars being sold through the dealer is equal to the total measure of high-valuation buyers. In this case, it is sufficient to assume that  $k > 0$  to rule out transactions of low-quality cars through the dealer. 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 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 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 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  $\phi U_t$ , 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  $U_t - p$ . 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

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

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 private market. Finally, it is easy to see that a low-quality seller will only sell in the private market. Hence, the above strategy profile is an equilibrium.

**Price Premium Dynamics.** In the market, both high-quality cars and low-quality cars are traded; the *average price* is therefore given by

$$b_t = \frac{(1 - \alpha)q_t}{1 - \alpha q_t} \phi U_t,$$

and we can compute the dealer's price premium in difference terms:

$$p_t - b_t = U_t^H \left[ \frac{1 - q_t}{1 - \alpha q_t} \right]$$

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

$$\frac{p_t}{b_t} = \frac{1/q_t - \alpha - (1 - \alpha)(1 - \phi)}{\phi(1 - \alpha)}. \quad (10)$$

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.

**Resales.** 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 private 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 private market, the resell rate is

$$\frac{(1 - \alpha)q_t + (1 - q_t)\delta}{1 - \alpha q_t}. \quad (11)$$

The intuition behind the previous formula is as follows. In total, the measure of cars being traded in the 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 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 previous section, the analysis is omitted.

### 3.4 Discussion

We now discuss some of the model’s features.

**Search Frictions.** Our theory based on consumer heterogeneity in Section 3.3 relies on the presence of implicitly assumed search frictions. The dealer’s role is to selectively save search cost for high-valuation buyers. Our setup is static and captures the search frictions in an extreme form: a buyer can only sample one seller despite the product heterogeneity in the market. Our reasoning remains valid as long as the search frictions exist. Suppose that in the market buyers and sellers randomly pairwise match in each period. Upon a match, the buyer observes the quality of the car, and the seller makes a take-it-or-leave-it offer. If a buyer rejects it, both agents go back to the market and are randomly matched in the next period. If sampling is costly, then by the insight of the Diamond paradox (Diamond 1971), in equilibrium, no offer will be rejected in the first period and the equilibrium transactions are identical to the ones in our model.

**Consumer Heterogeneity.** In the selection theory, we focus on the heterogeneity on consumers’ valuation, but other consumer heterogeneity may also help to explain the empirical observation. For example, one can introduce differential search/hassle costs to add more consumer heterogeneity, but to make sense of reallocation, differential preferences are still necessary, making the search cost heterogeneity non-essential in our model. Further data on consumer side is needed to identify the composition of consumer heterogeneity.

**Other Selection Mechanisms.** 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 selection 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 \(2014\)](#).

**Repairing.** Another channel for dealers is that they 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. Imagine that the dealer can repair a low-quality car: he incurs some cost  $c$  to make a low-quality car a high-quality one. To avoid the trivial case, we assume that  $c$  is sufficiently small. Therefore, the dealer purchases all cars, makes them high-quality due to the reputation concerns, and only sells high-quality cars. Then the proportion of high-quality cars in the market will be  $q_t$ . When information is asymmetric, the dealer's price is  $U_t$  and the market price is  $q_t U_t$ , so the price premium in difference is  $(1 - q_t)U_t$ , and the price premium in ratio is  $1/q_t$ . It is easy to see that they can also be consistent with empirical Facts 1 and 2. Similarly, the resale rate is higher than if the car is purchased from the market as it may be a lemon. By the same logic, one can derive similar implications when information is symmetric.

## 4 Asymmetric Information versus Consumer Heterogeneity

In Section 2, we document empirical facts about the used car market that suggest that dealers sell higher-quality cars compared to the direct market. In Section 3, we showed that these empirical regularities could be rationalized based on two leading theories about the role of intermediaries. The fundamental difference between our two theories lies in the assumption of consumers' information sets. 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).

### 4.1 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, i.e.,  $\dot{q}_t < 0$ . 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 it 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 private 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 market, and it is  $\delta$  for cars purchased from the dealer. Simple algebra implies that the resale rate gap between the market cars and dealer’s cars is

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

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 (11) for cars purchased from the market, and it remains  $\delta$  for cars purchased from the dealer. The resale rate gap between the private market cars and the dealer’s cars is

$$\frac{(1 - \alpha)(1 - \delta)}{\frac{1}{q_t} - \alpha}, \quad (13)$$

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 resale rate gap between the market cars and dealer cars.

Since these two tests require us to keep track of the transaction history of cars, we will use the Pennsylvania data; see section 2.2.1 for the data description. Figure 7 displays the resale rates of dealer sales and private sales for four time windows: one quarter, two quarters, three quarters, and four quarters. Private sales have higher resale rates for cars of all ages younger than 15 years, where after 15 years most cars are likely of low quality and we observe very few dealer sales. From ages 4 through 10, resale rates for private cars out-pace dealer cars. And then after 10 years, dealer car resale rates catch back up with private cars. The rates are similar for very old cars. The patterns are similar across different time windows.

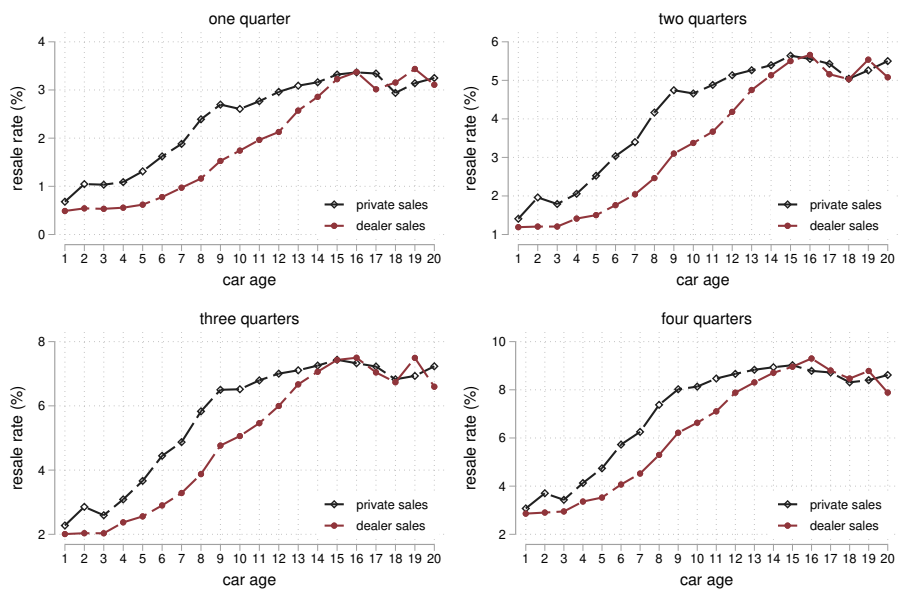


Figure 7: Resale Rates by Car Age

Note: The sample includes all unique used cars transacted from January 2014 to July 2015 in Pennsylvania. Sample selection is described in section 2.2.1. Black (red) lines represent the percentage of cars that were sold by private sellers (dealers) and were resold within a specific time window. Data source: Pennsylvania Department of Transportation.

### 4.1.1 Empirical Results

To make sure that these resale patterns are not driven by the composition of cars across car ages, we re-estimate equation (1), but add interactions of the car source and car age dummies. The detailed estimates of the coefficients for the age-private interactions are reported in Figure A.7 in the Appendix. For all specifications, the interaction coefficients increase monotonically until age eight and thereafter fall in car age. The estimates are noisy after age 13, mainly because the sample contains very few transactions of very old cars.

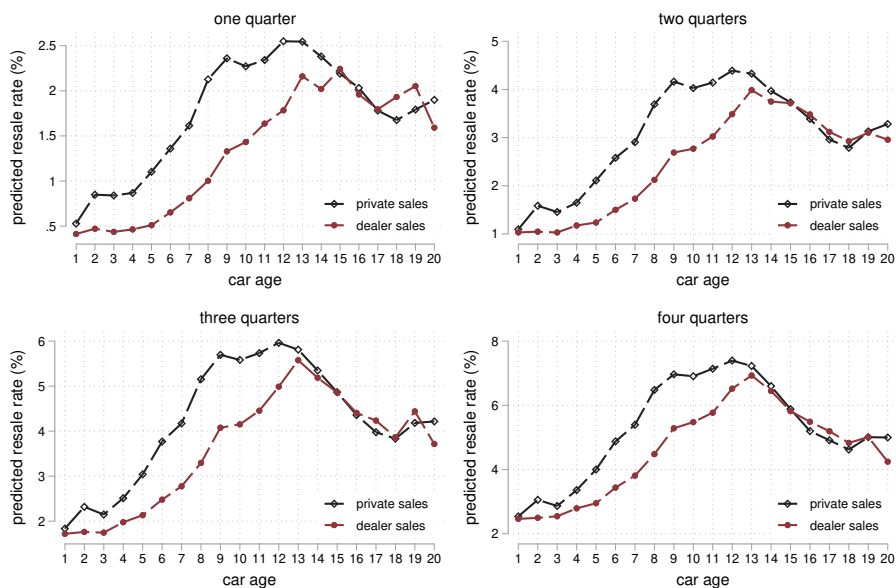


Figure 8: Predicted Resale Rates from FE Logit

Note: Predicted resale rates from FE Logit with age-source interactions.

Based on the estimates, we compute the predicted resale rates of cars and plot the predicted resale rates by age in Figure 8. Even after controlling for car attributes, the pattern is largely unchanged. Some interesting patterns emerge. First, the difference in resale rates is hump-shaped in the age of the car. Second, the percentage difference in the resale rates between dealer sales and private sales is the greatest for one-quarter resales. This is shown by the gap between the red and blue lines in Figure 8 which is greatest for one-quarter resales. For example, the predicted one-quarter resale rate for a 6-year-old privately sourced car is more than twice that of a car sourced from a dealer, but the predicted four-quarter resale rate for a 6-year-old privately sourced car is only about 50% greater than the resale rate of a similar car that was sourced from a dealer. Our

conclusion is that the information theory seems to dominate for younger cars (the diverging resale rates for cars younger than nine years old) and the sorting theory dominates for older cars (the converging and decreasing resale rates).

#### 4.1.2 Empirical Results Accounting for Endogeneity of Initial Purchase

As we did before, 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 2.2.3. 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 to 3, (ii) young cars ages 4 to 6, (iii) medium-age cars ages 7 to 10, (iv) old cars ages 11 to 15, and (v) very old cars beyond 15 years. Table 10 reports the number of sales, percentages of resales within one quarter and two quarters across the five age groups. The resale patterns in the restricted sample are similar to those of the entire sample, reported in Figure 7 in section 4.1.1, where private sales have higher resale rates than dealer sales over all age groups.

Table 10: Resale and Car Age

	Dealer Sales			Private 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 sample 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 2.2.3. Source: Pennsylvania Department of Transportation and Cars.com.

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 private seller, interactions of private 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 11 report the estimates of the interaction terms that capture the resale differences of private sales and dealer sales. The age-source interactions paint a similar picture as in Figure A.7 – the resale difference between private 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.

Table 11: Resale Difference between Private Sales and Dealer Sales

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

To address the endogeneity problem, we instrument the car source with the dealer inventory at the zipcode-style-week level, as in section 2.2.4. In the first stage, we run a regression of whether a car was bought from a private 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 private sellers. The second-stage estimates of the interaction terms are reported in the second and last columns of Table 11. The estimates of the coefficient before the indicator of very young cars (1-3 year) are not significant at 10% significant level. The estimates for very old cars are noisy, because there are both very few dealer and private sales. The estimates for 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 private 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, while dealers primarily promote sorting efficiencies as cars become very old.



## 4.2 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 would suggest 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; while with sorting, resale cars are more likely to be of high quality, so the equilibrium resale price will be higher.

As we described in section 2.2.1, 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.<sup>27</sup> For each resale, we calculate the difference between the resale price and its initial transaction price.

Figure 9a 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 one 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 9b and 9c we present the mean and median price changes in dollars and in percentage terms. When resales occur only one 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 duration gets longer, the difference between the resale price and the initial price falls. After two months, the median decrease is roughly 4%, and after three 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 market, given the higher proportion of price

---

<sup>27</sup>First, using the entire PA-DOT data that includes more than 2.3 million used car transactions registered in PA from January 2014 to July 2016, we estimate a hedonic regression of transaction price on various transaction characteristics including indicators for car ages, car age indicators interacted with seller type (dealer or private seller), log of mileage, and time effects, controlling for make-model-year-trim fixed effects. Then, we compute the predicted price based on the estimates. If a transaction price is four times higher or one quarter lower than the predicted price, we consider it as an outlier due to being mistakenly recorded or other reasons.

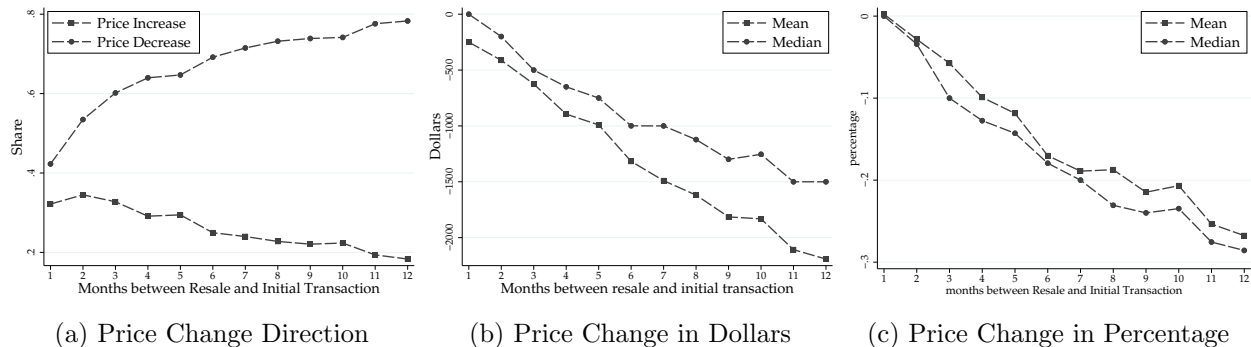


Figure 9: Resale Price and Initial Transaction Price

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

decreases and the negative median and average resale prices. We do not read too much into the price changes after three 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.

## 5 Conclusion

Although there is a rich theory literature that connects product intermediation to product quality (Biglaiser, 1993; Biglaiser and Friedman, 1994; Albano and Lizzeri, 2001; Bardhan, Mookherjee, and Tsumagari, 2013), 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 work 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.

## A Appendix: Additional Empirical Analysis

### A.1 Price Regressions with Seller County Fixed Effects

Here, we report results from the price premium regressions and the consumer reliability regressions that include seller county fixed effects. One thing to note is that we do not observe seller county for about one million observations. We display the summary statistics from this reduced sample of transactions in Table A.1. In Figure A.1 we display the predicted price premium for price regressions that include dummies for the seller’s county. The results are consistent with the baseline results in Figure 4.

Table A.1: Summary of Virginia DMV Data, Seller Location Sample

	Mean	SD	Q25	Q50	Q75
<b>Private Sales</b>					
Price	3,540	4,452	1,000	2,000	4,000
Mileage	137,590	65,031	96,720	135,421	174,245
Car Age	11.53	4.23	9	12	15
<b>Dealer Sales</b>					
Price	13,314	8,317	6,990	12,500	17,900
Mileage	75,586	51,496	35,621	64,511	105,384
Car Age	5.92	4.00	3	5	9
<b>Dealer Sales: 61.88%</b>					
<b>Total Transactions: 4,147,299</b>					

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

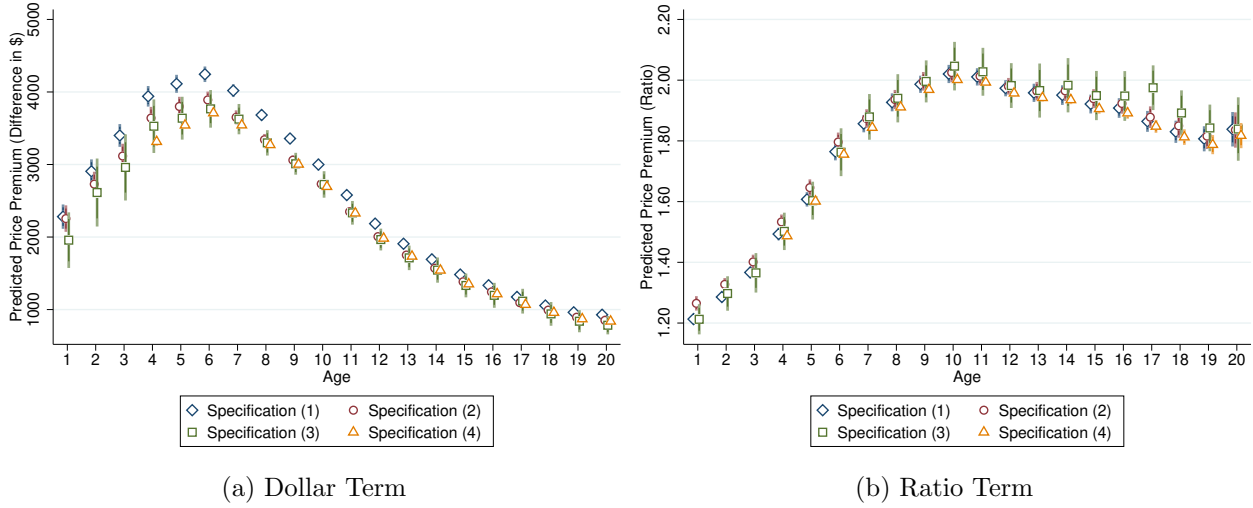


Figure A.1: Predicted Dealer Premium, with Seller County Fixed Effects

Note: Point estimates with 99% confidence intervals. Different specifications refer to the different columns in Table 3. Specification (1) includes 4,046,996 observations. Specification (2) excludes cars sold by new car dealers and includes 2,856,907 observations. Specification (3) includes only those popular products with 875,862 observations. Specification (4) includes cars older than three years old.

## A.2 Dealer Price Premium: Matching Estimator

Our OLS with product fixed effects exactly compares the prices of two otherwise observably identical (to the research) cars, where one is sold by a dealer and the other one is sold by a private seller. As a robustness, we also estimate the average treatment effect (ATE) of a dealer sale on price, by car age, using a matching estimator. To implement the matching estimator, we exactly match dealer and private cars on the following variables: make, model, trim, model year, mileage,<sup>28</sup> and seller county. To estimate the ATE of the ratio of dealer price over private price, we estimate the ATE of log-prices and then use the delta method to transform the standard errors from those for an estimate of the log difference to the ratio. Although the identification assumptions for the matching method is the same as our OLS method, the matching strategy does not rely on the linearity assumption of OLS.

We estimate the price premium ATE separately for each car age and display the 95% confidence intervals in Figure A.2. The overall patterns are very similar to the results using our OLS fixed effects estimator. The main difference is that the matching estimator has much higher standard errors, which is expected because a lot of the sample is lost from the exact matching procedure.

<sup>28</sup>We create a coarse bin for milages. In specific, we use bins of 30k miles as in Peterson and Schneider (2014).

In Table A.2, we display the coefficients, standard errors, and number of observations matched for each age.

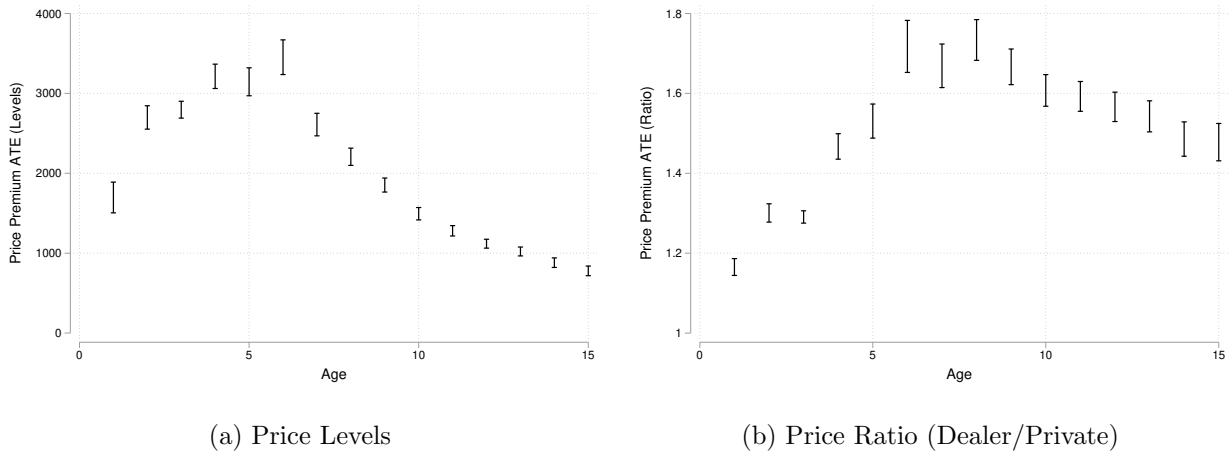


Figure A.2: Results of Matching Estimation

Note: We use the STATA package *teffects* to implement the exact matching estimator, and we compute standard errors according to Abadie and Imbens (2006). The outcome variable for "Levels" is car price. The outcome variable for "Ratio" is  $\log(\text{price})$  and we then transform the estimate using the exponential function and adjust the standard errors using the delta method.

Table A.2: Matching Estimates

Car Age	Level			Ratio		
	Coef.	SE	Obs	Coef.	SE	Obs
1	1,697.192	98.222	15,554	1.165	0.011	15,554
2	2,698.928	74.658	21,778	1.301	0.012	21,778
3	2,795.708	53.797	39,534	1.291	0.008	39,534
4	3,213.649	77.652	12,253	1.467	0.016	12,253
5	3,145.254	89.260	6,405	1.531	0.022	6,405
6	3,452.874	110.718	3,750	1.718	0.033	3,750
7	2,609.764	71.967	5,350	1.669	0.028	5,350
8	2,206.735	55.106	6,740	1.734	0.026	6,740
9	1,852.749	44.746	9,134	1.667	0.023	9,134
10	1,493.896	39.514	11,386	1.607	0.020	11,386
11	1,280.271	33.101	13,311	1.592	0.019	13,311
12	1,118.810	28.406	13,993	1.566	0.019	13,993
13	1,021.496	28.390	12,934	1.543	0.020	12,934
14	881.258	30.396	10,988	1.486	0.022	10,988
15	778.622	30.750	9,344	1.478	0.024	9,344

Note: We use the STATA package *teffects* to implement the exact matching estimator, and we compute standard errors according to Abadie and Imbens (2006). The outcome variable for "Levels" is car price. The outcome variable for "Ratio" is  $\log(\text{price})$  and we then transform the estimate using the exponential function and adjust the standard errors using the delta method.

### A.3 Dealer Price Premium and Car Reliability

Our main findings on price premium are that the dealer premium difference is hump-shaped in age and the dealer premium ratio is increasing in age. We perform an additional test to confirm that these results are driven by sellers' private information by merging our data with reliability ratings from *Consumer Reports* for 2005 through 2014.

*Consumer Reports* is a non-profit organization that publishes reviews of consumer products. The reliability ratings for automobiles come from surveys of car owners and expert testing and research.<sup>29</sup> The ratings are reported at the car model and model year level and range from 1 to 5, with a more reliable car model being rated with a higher number. We have ratings starting for the 2005 model year, and since our data ends in 2014 the oldest cars we have in the merged sample are 2005 model year cars sold as used cars in 2014. Also, not all make-models in our sample are rated by *Consumer Reports*. We report the summary statistics for the merged sample in Table A.3. Because of the merge with *Consumer Reports*, we are left with 1.3 million transactions that are younger, on average, than our original sample. The average rating in our sample is 3.39 for cars sold from dealers and 3.25 for cars sold from private sellers, so there doesn't appear to be

<sup>29</sup>We use an overall reliability rating for each car, which aggregates ratings for different car components. See Peterson and Schneider (2014) for further details of these data and a use of the component-level reliability ratings.

meaningful selection across seller type on the *Consumer Reports* rating. Roughly 86 percent of the transactions are sold from dealers, about 20 percentage points higher than our original sample.<sup>30</sup> We also report the highest- and lowest-rated cars among the top 50 percent of most popular cars in our sample. The cars listed are likely familiar as typically high and low reliability cars to those with a passing knowledge of the automobile industry.

Table A.3: Descriptive Statistics from Matched Consumer Reports Sample

	Mean	SD	Q25	Q50	Q75
<b>Private Sales</b>					
Price	10,999	7,327	5,300	10,000	15,000
Mileage	73,437	50,899	36,637	63,985	99,936
Car Age	4.63	2.33	3	4	7
Rating	3.39	1.20	3	3	4
<b>Dealer Sales</b>					
Price	17,137	6,978	12,528	15,998	20,858
Mileage	50,199	34,393	26,989	41,065	66,374
Car Age	3.31	1.99	2	3	4
Rating	3.25	1.19	3	3	4
<b>Dealer Sales: 85.88%</b>					
<b>Total Transactions: 1,362,195</b>					
<b>Most Reliable Models</b>			<b>Least Reliable Models</b>		
Honda CR-V			Honda Odyssey		
Toyota Corrolla			Ford Explorer		
Honda Accord			Chevrolet Impala		
Honda Civic			Chevrolet Cobalt		
Toyota Camry			Ford Escape		

Note: Summary statistics after merging baseline sample with *Consumer Reports* reliability ratings. Unit of observation is a single transaction. Car rankings are conditional on the car model being in the top 50 percent of sales in our sample.

We add the *Consumer Reports* reliability rating as an additional variable to the hedonic price regressions discussed in Section 2, Table 3. We interact the reliability rating with the seller type indicator, age indicators, and a triple interaction with both seller type and age. Our hypothesis is that car models that are more reliable have lower asymmetric information so there is less of a role for dealers to screen unobserved quality. Accordingly, we expect the age-slope of the dealer premium for more reliable cars to be less steep than less reliable cars. In other words, the dealer premium should not increase as much with age for reliable cars as for unreliable cars.

<sup>30</sup>We do a robustness by merging the average rating of a make-model with all model-years from the original sample. This leaves us with substantially more transactions and an age profile similar to the original sample. Our results are quantitatively very similar. However, by taking the average rating across model years we throw out important variation in rating within car model. For example, many Chevrolets go from low ratings to high ratings in 2011.

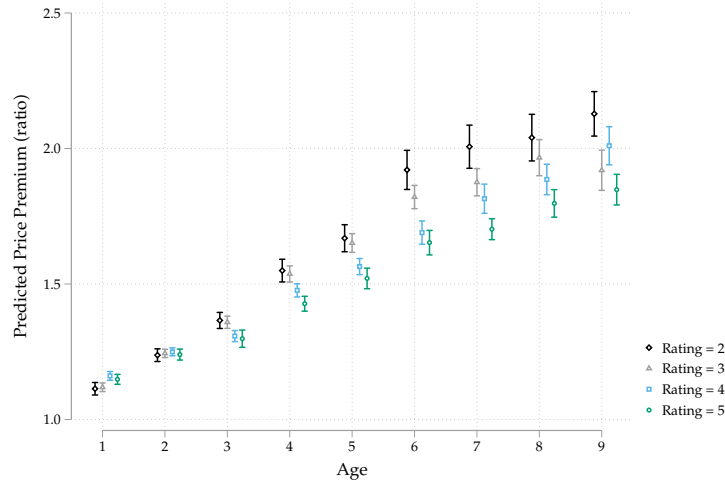


Figure A.3: Predicted Dealer Premium for Different Reliability Ratings

Note: Point estimates of the predicted price premium with 95% confidence intervals. The price premiums are computed by conditioning on reliability rating in the data. Premia predictions for cars with reliability equal to one are have very large associated standard errors across all ages due to very limited observations. Rating of “5” is the most reliable.

We display the predicted dealer premium ratios for different car reliabilities from the triple interaction regression described above in Figure A.3. The results correspond to the analogous hedonic price regression from specification (1) in Table 3, except with the additional reliability variable and corresponding interactions. We predict the dealer premium ratio at five values of reliability to give a sense of the gradient of the price premium with respect to reliability. The estimates suggest that less reliable cars have dealer premia that are steeper with respect to age. Not surprisingly, the price premium is nearly identical for all early vintage cars, regardless of reliability rating. This makes sense, as newer cars tend to have very low chances of suffering a defect in general, and the reliability ratings may be based off of defects that occur at later ages. But as car age increases, the premia across reliability ratings diverge.

**Fact 5.** *The dealer price premium for less reliable cars is more steeply increasing in age than that for more reliable cars.*

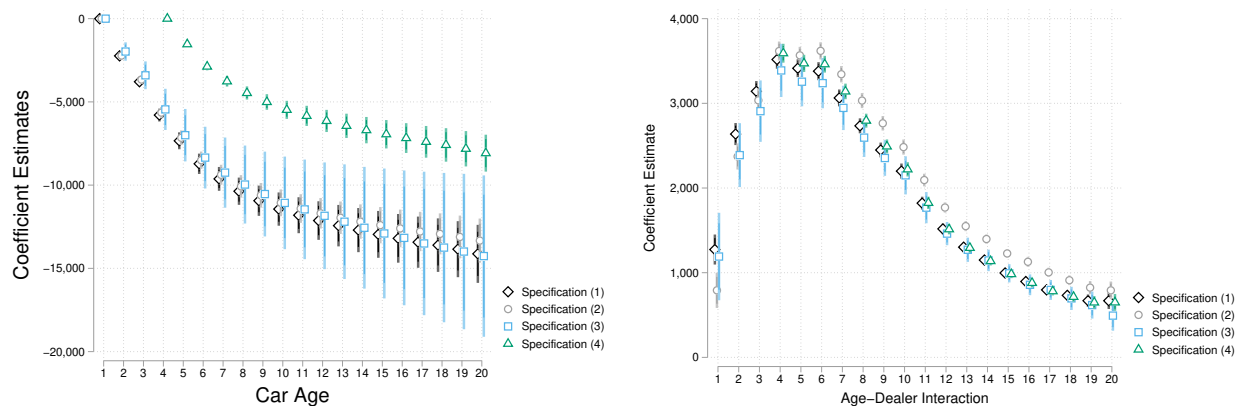


## A.4 Estimation Results of the Dealer Premium Equation with Price as the Dependent Variable

Table A.4: Dealer Premium Regressions

	(1)	(2)	(3)	(4)
log(Mileage)	-2,081	-1,962	-1,935	-2,257
Constant	39,512	37,514	36,291	27,183
Age Effects	..... See Figure A.4a .....			
Age-Dealer Interactions	..... See Figure A.4b .....			
$R^2$	0.582	0.558	0.645	0.529
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 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 seller indicator. All point estimates are statistically significant at least at the 0.001 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 four years old.



(a) Car Age Dummies

(b) Car Age-Dealer Interactions

Figure A.4: Coefficient Estimates

Note: Point estimates with 99% confidence intervals. Different specifications refer to the different columns in Table 3.

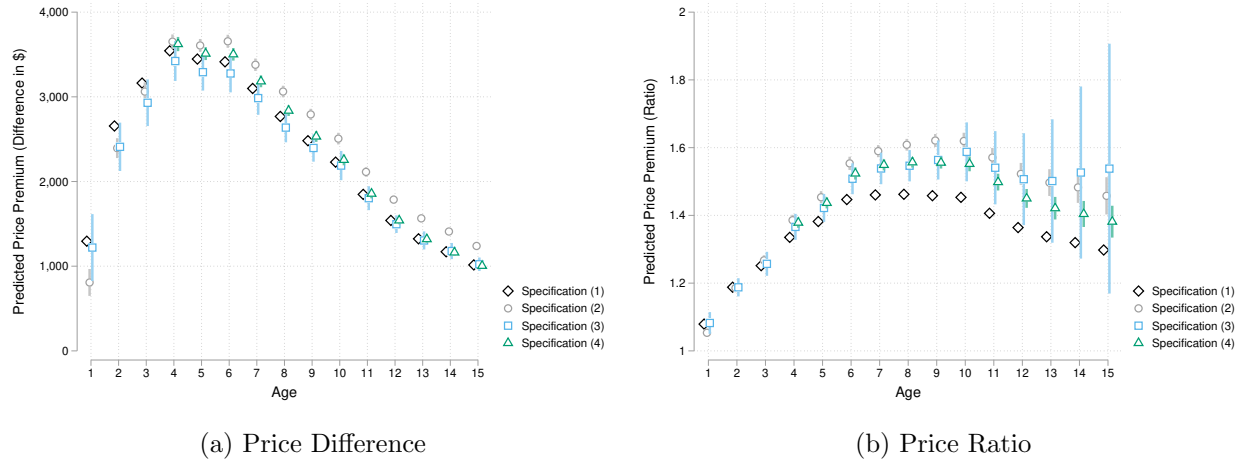


Figure A.5: Predicted Dealer Premium

Note: Point estimates with 99% confidence intervals. Different specifications refer to the different columns in Table A.4.

## A.5 Estimation Results of the Resale Analysis with the Full Sample

Table A.5: Immediate Resale and Car Source: Logit with Product Fixed Effects <sup>[H]</sup>

	Resale Time Window			
	One Quarter	Two Quarters	Three Quarters	Four Quarters
Bought from Dealer	-0.392 (0.018)	-0.259 (0.013)	-0.186 (0.011)	-0.144 (0.009)
Log Mileage	0.135 (0.019)	0.179 (0.014)	0.176 (0.011)	0.170 (0.010)

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 sample includes 1,430,307 unique used cars transacted from January 1, 2014, to July 31, 2015, in Pennsylvania.

Table A.6: Immediate Resale and Car Source: Logit with Control Function

	Fixed Effects Logit		Control Function	
	Resale Window		Resale Window	
	One Quarter	Two Quarters	One Quarter	Two Quarters
Bought from Dealer	-0.450 (0.069)	-0.410 (0.051)	-0.465 (0.075)	-0.445 (0.055)
Log Mileage	0.244 (0.079)	0.289 (0.058)	0.167 (0.188)	0.424 (0.124)

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. In the control function panel, we use dealer inventory as the excluded variable for whether a car was bought from a dealer. Standard errors in parentheses. The sample includes 73,803 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.

Table A.7: Robustness Checks of the Control Function Approach

	Linear Probability		Alternative IV	
	Resale Window		Resale Window	
	One Quarter	Two Quarters	One Quarter	Two Quarters
Bought from Dealer	-0.460 (0.075)	-0.447 (0.055)	-0.457 (0.069)	-0.408 (0.051)
Log Mileage	0.263 (0.088)	0.311 (0.064)	0.551 (0.180)	0.587 (0.134)

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. In the linear probability panel, we use a linear probability model with product fixed effects to model the first stage and use the dealer inventory as the excluded variable for whether a car was bought from a dealer. In the alternative IV panel, we use a Logit model with product fixed effects to model the first stage, but use the log of the distance to the nearest CarMax as the excluded variable for whether a car was bought from a dealer. Standard errors in parentheses. The sample includes 73,803 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.

Table A.8: Resale Difference between Private Sales and Dealer Sales

	One Quarter Resale		Two Quarter Resale	
	OLS	IV	OLS	IV
(i) Very Young	0.002 (0.002)	-0.202 (0.164)	0.005 (0.003)	0.023 (0.221)
(ii) Young	0.005 (0.003)	0.037 (0.018)	0.005 (0.005)	0.079 (0.024)
(iii) Medium	0.010 (0.003)	0.058 (0.018)	0.015 (0.004)	0.088 (0.024)
(iv) Old	0.006 (0.003)	0.070 (0.020)	0.008 (0.004)	0.096 (0.027)
(v) Very Old	0.004 (0.006)	0.032 (0.037)	0.005 (0.008)	0.039 (0.050)

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 sample includes 73,803 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.

## A.6 Additional Figures and Tables

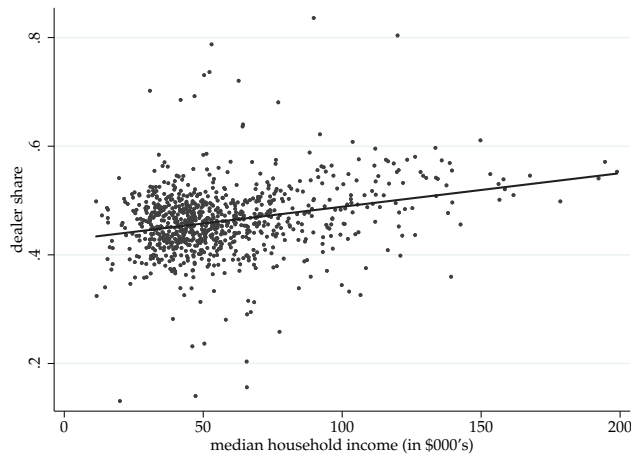


Figure A.6: Dealer Share and Median Household Income

Note: An observation is a zipcode in Virginia. Dealer share refers to the proportion of dealer sales among all used-car sales that were bought by buyers in the same zip code. Sample includes all used-car transactions that were registered in Virginia DMV from 2007 to 2014.

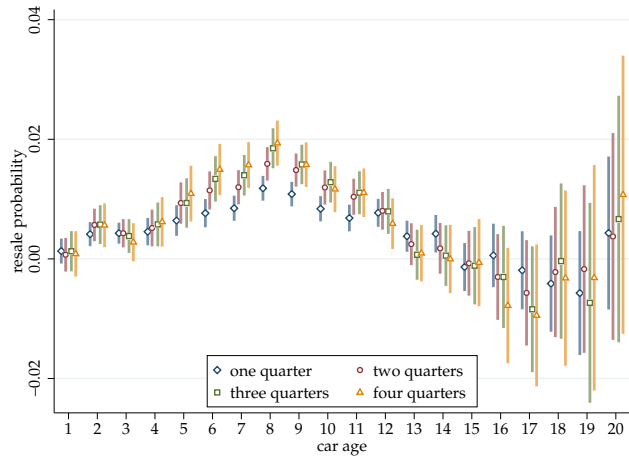


Figure A.7: Car Age-Private Interactions

Note: Point estimates with 99 % confidence interval.

## References

- ABADIE, A., AND G. W. IMBENS (2006): “Large sample properties of matching estimators for average treatment effects,” *econometrica*, 74(1), 235–267.
- ADAMS, C., L. HOSKEN, AND P. NEWBERRY (2011): “Vettes and Lemons on eBay,” *Quantitative Marketing and Economics*, 9, 109–127.
- ADAMS, W., L. EINAV, AND J. LEVIN (2009): “Liquidity constraints and imperfect information in subprime lending,” *The American Economic Review*, 99(1), 49–84.
- AKERLOF, G. A. (1970): “The market for” lemons”: Quality uncertainty and the market mechanism,” *The Quarterly Journal of Economics*, pp. 488–500.
- ALBANO, G. L., AND A. LIZZERI (2001): “Strategic certification and provision of quality,” *International economic review*, 42(1), 267–283.
- BARDHAN, P., D. MOOKHERJEE, AND M. TSUMAGARI (2013): “Middlemen margins and globalization,” *American Economic Journal: Microeconomics*, 5(4), 81–119.
- BERNDT, A., AND A. GUPTA (2009): “Moral hazard and adverse selection in the originate-to-distribute model of bank credit,” *Journal of Monetary Economics*, 56(5), 725–743.
- BIGLAISER, G. (1993): “Middlemen as experts,” *The RAND journal of Economics*, pp. 212–223.
- BIGLAISER, G., AND J. W. FRIEDMAN (1994): “Middlemen as guarantors of quality,” *International journal of industrial organization*, 12(4), 509–531.
- (1999): “Adverse selection with competitive inspection,” *Journal of Economics & Management Strategy*, 8(1), 1–32.
- BIGLAISER, G., AND F. LI (2018): “Middlemen: the good, the bad, and the ugly,” *The RAND Journal of Economics*, 49(1), 3–22.
- BOND, E. W. (1982): “A Direct Test of the ”Lemons” Model: The Market for Used Pickup Trucks,” *American Economic Review*, 72, 836–840.
- BOND, E. W. (1983): “Trade in used equipment with heterogeneous firms,” *Journal of Political Economy*, 91(4), 688–705.
- BOND, E. W. (1984): “Test of the Lemons Model: Reply,” *American Economic Review*, 74, 801–804.

- CARDON, J. H., AND I. HENDEL (2001): “Asymmetric information in health insurance: evidence from the National Medical Expenditure Survey,” *RAND Journal of Economics*, pp. 408–427.
- DIAMOND, P. A. (1971): “A model of price adjustment,” *Journal of economic theory*, 3(2), 156–168.
- DONNA, J., A. TRINDADE, P. PEREIRA, T. PIRES, ET AL. (2018): “Measuring the Welfare of Intermediation in Vertical Markets,” in *2018 Meeting Papers*, no. 984. Society for Economic Dynamics.
- EINAV, L., A. FINKELSTEIN, AND P. SCHRIMPF (2010): “Optimal mandates and the welfare cost of asymmetric information: Evidence from the uk annuity market,” *Econometrica*, 78(3), 1031–1092.
- ENGERS, M., M. HARTMAN, AND S. STERN (2009): “Are Lemons Really Hot Potatoes?,” *International Journal of Industrial Organization*, 27, 250–263.
- GALENIANOS, M., AND A. GAVAZZA (2017): “A Structural Model of the Retail Market for Illicit Drugs,” *American Economic Review*, 107(3), 858–96.
- GALENIANOS, M., R. L. PACULA, AND N. PERSICO (2012): “A search-theoretic model of the retail market for illicit drugs,” *The Review of Economic Studies*, 79(3), 1239–1269.
- GAVAZZA, A. (2016): “An empirical equilibrium model of a decentralized asset market,” *Econometrica*, 84(5), 1755–1798.
- GEHRIG, T. (1993): “Intermediation in search markets,” *Journal of Economics & Management Strategy*, 2(1), 97–120.
- GENESOVE, D. (1993): “Adverse selection in the wholesale used car market,” *Journal of Political Economy*, pp. 644–665.
- GUERRIERI, V., AND R. SHIMER (2014): “Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality,” *American Economic Review*, 104(7), 1875–1908.
- HENDEL, I., AND A. LIZZERI (1999): “Adverse Selection in Durable Goods Markets,” *American Economic Review*, 89(5), 1097–1115.
- HENDEL, I., A. NEVO, AND F. ORTALO-MAGNÉ (2009): “The relative performance of real estate marketing platforms: MLS versus FSBO Madison.com,” *The American Economic Review*, 99(5), 1878–1898.

- HENDREN, N. (2013): “Private information and insurance rejections,” *Econometrica*, 81(5), 1713–1762.
- HWANG, I. (2018): “Dynamic trading with developing adverse selection,” *Journal of Economic Theory*, 176, 761–802.
- JOHRI, A., AND J. LEACH (2002): “Middlemen and the allocation of heterogeneous goods,” *International Economic Review*, 43(2), 347–361.
- KEYS, B. J., T. MUKHERJEE, A. SERU, AND V. VIG (2010): “Did securitization lead to lax screening? Evidence from subprime loans,” *The Quarterly journal of economics*, 125(1), 307–362.
- KIM, K. (2012): “Endogenous market segmentation for lemons,” *The RAND Journal of Economics*, 43(3), 562–576.
- LACKO, J. M. (1986): “Product Quality and Information in the Used Car Market,” *Bureau of Economics Staff Report to the Federal Trade Commission*.
- LARSEN, B. (2014): “The efficiency of real-world bargaining: Evidence from wholesale used-auto auctions,” Discussion paper, National Bureau of Economic Research.
- LESLIE, P., AND A. SORENSEN (2013): “Resale and rent-seeking: An application to ticket markets,” *Review of Economic Studies*, 81(1), 266–300.
- LIZZERI, A. (1999): “Information revelation and certification intermediaries,” *The RAND Journal of Economics*, pp. 214–231.
- MIAN, A., AND A. SUFI (2009): “The consequences of mortgage credit expansion: Evidence from the US mortgage default crisis,” *The Quarterly Journal of Economics*, 124(4), 1449–1496.
- MURRY, C., AND H. SCHNEIDER (2015): “The Economics of Retail Markets for New and Used Cars,” in *Handbook on the Economics of Retail and Distribution*, ed. by E. Basker. Edward Elgar.
- NOSAL, E., Y.-Y. WONG, AND R. WRIGHT (2015): “More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets,” *Journal of Money, Credit and Banking*, 47(S2), 7–37.
- (2017): “Intermediation in Markets for Goods and Markets for Assets,” *Working Paper, University of Wisconsin Madison*.



- PETERSON, J. R., AND H. S. SCHNEIDER (2014): “Adverse selection in the used-car market: evidence from purchase and repair patterns in the Consumer Expenditure Survey,” *The RAND Journal of Economics*, 45(1), 140–154.
- RHODES, A., M. WATANABE, AND J. ZHOU (2018): “Multiproduct Intermediaries,” Discussion paper, Yale University.
- RUBINSTEIN, A., AND A. WOLINSKY (1987): “Middlemen,” *The Quarterly Journal of Economics*, 102(3), 581–593.
- RUST, J., AND G. HALL (2003): “Middlemen versus market makers: A theory of competitive exchange,” *Journal of Political Economy*, 111(2), 353–403.
- SALZ, T. (2017): “Intermediation and Competition in Search Markets: An Empirical Case Study,” *working paper, Columbia University*.
- SHEVCHENKO, A. (2004): “Middlemen,” *International Economic Review*, 45(1), 1–24.
- SPULBER, D. F. (1996): “Market Making by Price-Setting Firm,” *The Review of Economic Studies*, 63(4), 559–580.
- WRIGHT, R., AND Y.-Y. WONG (2014): “Buyers, Sellers, And Middlemen: Variations On Search-Theoretic Themes,” *International Economic Review*, 55(2), 375–397.
- YAVAŞ, A. (1994): “Middlemen in bilateral search markets,” *Journal of Labor Economics*, 12(3), 406–429.
- (1996): “Search and trading in intermediated markets,” *Journal of Economics & Management Strategy*, 5(2), 195–216.