Automobile Maintenance Costs, Used Cars, and Private Information

Maxim Engers
University of Virginia

Monica Hartmann
University of St. Thomas

and Steven Stern
University of Virginia

November 16, 2012

1 Introduction

Since Akerlof's (1970) foundational work on lemons markets, economists have been investigating how differences in information about product quality negatively affect the amount of trade in durable goods. Cardon and Hendel (2001), Hendel, Lizzeri, and Siniscalchi (2005), and Engers, Hartmann, and Stern (2009a) (EHSa) among others, find that, in a dynamic setting with sufficient heterogeneity in tastes, the trade-inhibiting effects of private information are less severe than previously thought.

There is also an extensive literature on the role of maintenance expenditures in decisions to buy or sell durable goods. (See, for example, Rust (1987), Rust and Rothwell (1995), Cho and Rust (2008), and Bensoussan et al. (2011).) These models, however, do not allow for asymmetric information about maintenance to affect other decisions. In this paper, we examine the role that unobserved maintenance expenditures play in household selling decisions for durable goods. We identify private information through the link between the household maintenance and selling decisions not accounted for by household demographics and car characteristics.

The Bureau of Labor Statistics' Consumer Expenditure Survey tracks households' expenditures on vehicles. This dataset contains information on sales by vehicle owners and on the total cost of maintaining the household's portfolio of vehicles. The richness of the data potentially can help us identify and quantify private information in the used car market. However, we show that, when there is both private information and heterogeneity observed by the market but not by the econometrician, there is a fundamental identification problem. In particular, there are three parameters of interest, a) the amount of private information, b) the degree to which costs influence selling decisions, and c) heterogeneity in household selling preferences independent of vehicle characteristics, that have identifying information captured in only two moments of the data. Thus, we can identify only two of the three as a function of the other. We also propose an alternative model excluding the existence of any private information that also is consistent with the data. Thus, while we suggest that the data are consistent with the existence of a significant amount of private information, especially when one does not allow for unobserved heterogeneity observed by the market, they are also consistent with alternative models that look quite different.

EHSa provides evidence suggesting that unobservable heterogeneity in vehicle quality that drives selling decisions is not specific to a vehicle. Rather it varies across owners of the same vehicle. A possible explanation for this finding hinges on differences in tastes among new-car buyers: New-vehicle owners whose utility falls rapidly with declines in quality maintain their vehicles well and yet sell them soon after buying them. On the other hand, new-vehicle owners whose utility does not fall rapidly with declines in quality hold them for a long period and maintain them poorly. These findings suggest that the lemons problem may be due to individual owners' rather than manufacturers' decisions. In this paper, we show that one of the avenues through which this can happen is maintenance expenditures.

We model both the household maintenance decision and selling decision as a function of the vehicle's brand, age, interaction of these terms, and household demographics, such as number of drivers and household vehicles owned, and error terms. These variables capture differences in driving intensity and thus demand for a vehicle. The empirical analysis uses data from the Consumer Expenditure Survey which tracks households' expenditures on vehicles.[1] For each vehicle, we observe the make, model, and age. Total expenditures on maintenance for all vehicles owned by the household are given. The survey also provides information on the household's income, whether the household lives in an urban setting, and the number of drivers per vehicle in the household. This demographic information allows us to control for differences in vehicle operating costs that vary by household characteristics. Finally, we observe when vehicles are purchased or sold (or scrapped).

Our results indicate that an unobserved component in the cost of maintaining vehicles exists and influences household selling decisions as commonly assumed.[2] In fact, we find evidence that the cost threshold associated with the selling decision increases with prior unexpected costs of maintaining a vehicle, suggesting that households who intend to sell spend less on maintenance the period prior to selling. Thus, the individual owners' maintenance decisions are a source of the private information problem observed in the used vehicle market. In fact, we determine that this problem worsens as vehicles age, peaking at age 9.

In addition to demonstrating that private information may be important in

[1] We control for selection bias due to a sample construction rule.

[2] Both Fabel and Lehmann (2002) and Engers, Hartmann, Stern (2009b) define quality as a function of the owner's maintenance effort that is unobserved by potential buyers. This modeling of quality as owner-determined is consistent with optimal asset durability and retirement/scrappage literatures. See, for example, Schmalensee (1974), Parks (1979), Rust (1985), and Nelson and Caputo (1997).

the sales decision, the detailed survey of household expenditures on passenger vehicles permits us to create operating-cost profiles for automobiles by brand and age. We find evidence that the cost of maintaining a vehicle rises in the first 5 years, then levels off and declines after age 10. This is consistent with Pickrell and Schimek (1999) and Engers, Hartmann, and Stern (2009b) (EHSb). These studies observe significant reductions in vehicle usage after age 10, consistent with lower maintenance expenditures in later years. Knowledge of the operating cost profiles is of value for those studying automobile-related issues such as scrapping or buying and selling decisions where households compare the vehicle's sale (scrap) price against the cost of maintaining it.

The layout of the paper is as follows: Section 2 summarizes the literature on the selling and scrapping of durable goods. Section 3 describes the data used to estimate the cost of maintaining a vehicle as a function of its product characteristics as well as the owner's demographics. Section 4 presents the basic model of household expenditures on maintaining their vehicles and how these influence its vehicle selling decisions. Section 5 describes a selection-bias problem associated with the data and how we address it, Section 6 introduces private information into the model, and Section 7 addresses identification issues, especially those associated with the inclusion of private information. Section 8 discusses the results and their implications for the role maintenance expenditures play in selling decisions, and Section 9 analyzes the relationship between annual miles driven and maintenance cost across vehicle age to obtain a deeper understanding of maintenance decisions. Section 10 provides an alternative interpretation of our results.

2 Literature Review

To understand the market for used vehicles, one must understand household decisions to buy and sell durable goods. These goods are consumed over many periods and are often resold. Used goods are imperfect substitutes for new ones; their qualities depreciate over time and at different rates. Consumers' willingness to keep these goods as they depreciate varies with consumer preferences for quality.

The literature on buying and selling of durable goods can be classified into two strands based on assumptions made about the way information is distributed in the market. The first strand of the literature assumes complete information is available to buyers and sellers. These papers focus on the role quality deterioration plays in households' purchasing decisions. Rust (1985) provides a general theoretical model of stochastic degradation. It endogenizes the selling decision and allows for heterogeneous preferences for product quality. In stationary equilibrium, consumers hold durable goods for one period and repeatedly purchase the same quality grade. Konishi and Sandfort (2002) introduces a transaction cost for replacing a good. In equilibrium, there is an incentive for consumers to continue to hold durable goods as their quality changes. More recently, Esteban and Shum (2007) and Schiraldi (2011) specify and estimate a dynamic structural model of these replacement decisions. They take into account how durability and transaction costs affect dynamic consumer considerations in their vehicle replacement decisions.

In contrast, the second strand of the literature focuses on the effects of differences in information between owners and potential buyers. Akerlof (1970) argues that adverse selection may entirely shut down the market for second-hand durables. Jansen and Roy (2002) examines the effect of allowing sellers of durable goods in a lemons market to defer trade until later periods. Although this leads to more trade than in a static model, the private information leads to a socially inefficient amount of waiting: owners of higher-quality goods wait longer to trade, and both the price and the average quality of goods traded rise over time. When Hendel and Lizzeri (1999, 2002) extend Akerlof's model to allow for unobserved heterogeneity with infinite support, they conclude that asymmetric information reduces the amount of trade in equilibrium but never shuts it down. Hendel and Lizzeri (1999) constructs a test for adverse selection. Their results based on a theoretical model implies that durable goods markets suffer from adverse selection when the goods with the largest price declines are also the ones that are least traded in the secondary market. Used car market data are found to be consistent with adverse selection being present.

Johnson and Waldman (2010) presents a model in which both adverse selection and moral hazard are present. The paper focuses on the consumers' choice of leasing or buying a new car or buying a used one. Cars last only two periods and the quality of new cars is given and known. The quality of used cars is variable and privately known only to the first-period owner or leaser. It is determined both by a private stochastic shock (the adverse selection part) and the amount of costly maintenance undertaken (the moral hazard part). Owners who intend to sell will cut maintenance to the minimum, but leasing contracts can ameliorate the moral hazard problem by imposing minimum maintenance requirements.

This paper belongs to the second strand of the literature. We examine the role unobserved maintenance expenditures play in selling decisions and determine if they are the source of the asymmetric information problem previously found in the used car market.[3] There is an extensive literature that shows how maintenance expenditures affect the depreciation rate for durable goods and thus, the decision to replace them. See, for example, Rust (1987), Rothwell and Rust (1997), Cho and Rust (2008), Bensoussan et al. (2009), and Chen (2011). However, these studies assume there is no asymmetric information with respect to maintenance.

In our paper, we first estimate how much households spend on maintaining their vehicles. Maintenance expenditures are modeled as a function of the household's demographics, the vehicle's brand and age, the interaction of these terms, and error terms. A household sells (or scraps) a vehicle when its costs exceed a certain threshold. If the unobserved components of the cost of main-

[3] See, for example, Johnson and Waldman (2003), Hendel and Lizzeri (1999), and Emons and Sheldon(2009).

taining a vehicle influence selling decisions, then one can conclude that the used car market suffers from asymmetric information problems.

This paper contributes to the literature in two other ways as well. First, Cho and Rust (2008) finds no evidence that maintenance costs in the rental car market rise with odometer rating (or age, given that age and total miles driven are highly correlated). As a result, it assumes an average maintenance cost per day. While their method accurately predicts maintenance decisions for auto rentals, it is not reflective of household maintenance decisions for cars the household owns rather than rents. For example, household vehicle expenditures are more sporadic in nature. In addition, rental cars are used more intensively (i.e., higher annual miles driven) and are much younger than the typical car on the road. The maximum age in the Cho and Rust (2008) sample is 4.5 years which is much lower than the average age of a passenger vehicle on U.S. roads. We contribute to the literature then by estimating cost profiles over a vehicle's life by brand.

Our second contribution is through our clarification of the role maintenance costs play in how prices of durable goods change over the asset's life. Rust (1985) and Hendel and Lizzeri (1999) demonstrate theoretically how quality depreciation and/or asymmetric information reduce consumers' willingness to pay and thus decrease market price as a vehicle ages. In this paper, we analyze how household maintenance decisions vary by vehicle age, by household demographics that affect usage, and by the amount of asymmetric information present. A better understanding of these maintenance decisions provides insights on how a vehicle's quality depreciates endogenously over its lifetime. This augments researchers' understanding of what explains observed price declines over a vehicle's life.

3 Data

The Consumer Expenditure Survey (CEX) tracks household spending on various items including automobile-related expenses. It is a rotating panel in which each household (a family or a single consumer) is interviewed at most four times. Every quarter, 25% of the households in the sample are replaced with new households.

For each household, the survey records total expenditures on maintenance, vehicle insurance, gasoline and oil, licensing and registration fees, and other miscellaneous expenses. These combined household expenses reflect the total costs of operating a vehicle; thus, for the duration of the paper, we refer to the combined expenses as the costs of maintaining a vehicle. The survey also collects information on the stock of vehicles owned by the household, providing detailed information on each vehicle's make, model, and year. All of the information is collected quarterly.

Because expenditures are aggregated over a household's stock of vehicles, we place two restrictions on the sample. First, since we are focusing on passenger vehicles, we exclude quarterly observations for a household from our sample if

Table 1: Observed Cars By Brand

Brand	# Obs	Brand	# Obs	Brand	# Obs
Buick	5993	Mazda	2451	Toyota	7871
Chevrolet	11125	Mercury	3798	Volkswagen	1307
Chrysler	2438	Mitsubishi	1337	Volvo	942
Dodge	4446	Nissan	4752	American Luxury	3534
Ford	13265	Oldsmobile	5947	European Luxury	1669
Geo	1174	Plymouth	2464	Japanese Luxury	397
Honda	9146	Pontiac	5992	Other	143
Hyundai	1320	Saturn	1081	Truck	33010
		Subaru	1342		

Notes: 1. American Luxury includes Cadillac and Lincoln.
2. European Luxury includes Audi, BMW, Jaguar, Mercedes Benz, Porsche, and Saab.
3. Japanese Luxury includes Infinity and Lexus.
4. Other includes Isuzu, Kia, and Peugeot.

Table 2: Missing Observation Analysis

Reason	Households* Dropped	Interviews* Dropped	Households Remaining	Interviews Remaining
Initial Size			60889	151045
Region Missing	3268	9140	57621	141905
Income Missing	10400	15778	47221	126127
Vehicle Missing	15	15	47206	126112

* Recall households are interviewed more than once in the CEX dataset and those household interviews that do not match our sample criteria are dropped.

it owns a boat, motorcycle, or trailer in that quarter. Second, for our cost regressions only, we exclude household-quarters in which the household bought or sold a vehicle in that quarter. This is because the CEX does not indicate at what date the household incurs the expenses. As a result, we are unable to decompose household expenses when the household's stock of vehicles is changing within a quarter.

We put additional restrictions on the data set to meet our estimation needs. First, a quarterly observation for a household is included only if each of the household's passenger vehicles is manufactured by one of the auto makers listed in Table 1. Some of the manufacturer brand names have been combined for simplicity.[4] Second, we include only vehicles manufactured after 1985 because the CEX provides model-year information for only these vehicles. Table 2 provides information on the number of observations lost for each selection criterion. Given these restrictions, the sample consists of 126112 interviews of 47206 households between 1988 and 1999. There are 87140 unique vehicles

[4] Following EHSa and EHSb, we aggregate Cadillac and Lincoln into "American Luxury," Infiniti and Lexus into "Japanese Luxury," and Audi, BMW, Jaguar, Mercedes-Benz, Porsche, and Saab into "European Luxury." Honda and Acura vehicles are combined into a single brand category, as are Mitsubishi and Eagle vehicles. Rarer brands like Peugeot, Isuzu, and Kia are combined to form the "Other" category.

Table 3: Moments for CEX Data

Variable	# Obs	Mean	Std Dev
Maintenance Cost ($1000s)			
Total Real Expenditures	126112	0.698	0.603
Maintenance	126112	0.185	0.404
Insurance	126112	0.212	0.213
License & Registration	126112	0.029	0.083
Miscellaneous	126112	0.272	0.238
Household Demographics			
Urban	47206	0.958	0.201
Log Income	47206	10.295	0.932
# of Drivers	47206	1.879	0.870
# of Drivers/Car	47206	1.393	0.867
# of Cars Owned	126112	1.761	0.897
Vehicle Characteristic			
Age (at first interview)	45208	5.802	3.201

Note: except for number of cars owned, household demographics are based on information given at first interview.

owned by households over the sample period. The sample contains 218254 vehicles with most vehicles appearing multiple times. Table 3 reports the sample moments of the data. In our sample, 46.0% of households own 1 vehicle, 33.6% own 2 vehicles, and 20.4% own more than 2 vehicles; 0.7% of households own 6 or more vehicles. Quarterly household expenditures on vehicle-related items vary from as little as $0 to as high as $21410 (in 1999 dollars).

Table 4 shows how the median age of automobiles varies by calendar year for our sample versus the true population. Notice that, in the earlier years, the sample's median is much lower than for the population. We drop from the sample any household that owns a vehicle produced before 1986 because we lack price data for these model-year.[5] Later in the paper, we discuss how we correct for this selection-bias problem.

Figure 1 provides additional information on the entire age density of vehicles included in the sample at the initial interview. Most of the vehicles are between 2 and 8 years old. Figure 2 presents the density of age by brand for several representative brands. The age density is quite similar across brands except for Japanese Luxury which has a lower median age.[6]

We also observe some demographic information for each household. Eighty-eight percent of the households live in an urban area, with the Midwest region having the highest percentage of rural households. The average family size varies

[5] The distribution of vehicle age differs for another reason. The sample of households in the CEX is not representative of the U.S. without using population weights. Because we do not use these weights, we do not measure population correlation correctly between household characteristics, such as income and family size, and the true distribution of vehicle age.

[6] Acura and Lexus began to sell vehicles the US starting in 1986 and 1989, respectively. Because sales tend to be lower in the first years of entry, their median age is lower than the more established brands. Thus, Japanese Luxury has a younger age density.

Table 4: Median Age of Automobiles

Year	Population*	Sample
1990	6.5	5.6
1991	6.7	6.3
1992	7.0	7.1
1993	7.3	7.2
1994	7.5	7.5
1995	7.7	7.8
1996	7.9	8.0
1997	8.1	8.3
1998	8.3	8.2
1999	8.3	8.0

* Source: Table 1-22 Median Age of Automobiles in Operation in the U.S., National Transportation Statistics 2000

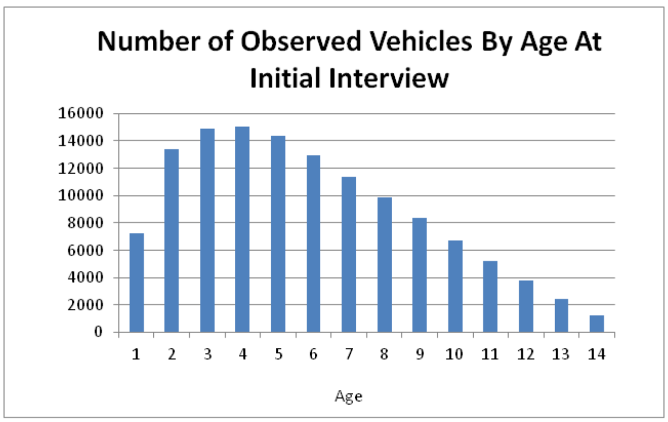

Figure 1: Number of Observed Vehicles By Age

across the country with the Midwest having the smallest average (2.51), while the West has the largest (2.61). The average household has 1.88 drivers with 17.3% having at least 3 drivers in the household. As seen in Figure 3, 71.0% of the households have 1 or fewer drivers per vehicle owned, and 4.0% have at least 3 drivers per vehicle. The variation in the number of drivers per household vehicle may explain the variation observed in maintenance expenses across households. The household's income (in 1999 dollars) is the final demographic variable observed.[7] The average income for households in the sample is $42221. Household income is observed to be as high as $123301 and as low as $1652.

We also control for other factors that affect the amount households spend

[7]The observed income measure is bracketed. Following Stern (1991), we match these discrete data with continuous income information from the Current Population Survey. From these survey data, we construct the mean income for each income bracket in every year of the sample.

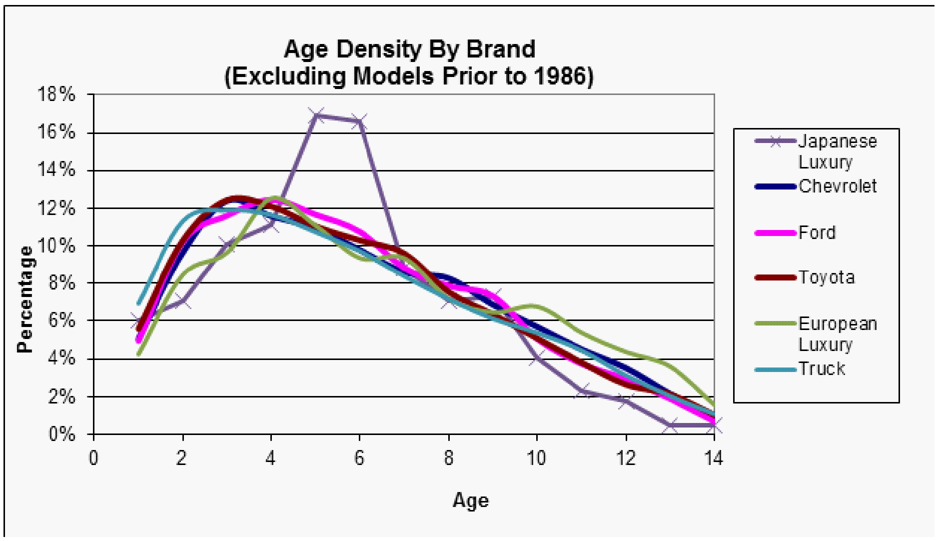

Figure 2: Age Density By Brand

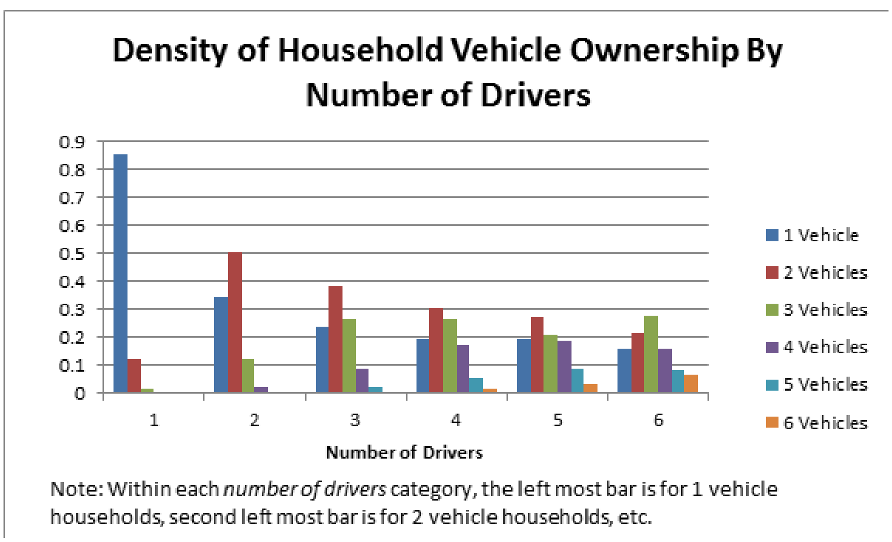

Figure 3: Density of Household Vehicle Ownership

on their vehicles beyond the household and vehicle characteristics in the CEX data set. For example, regional differences in driving patterns will affect the extent to which a household maintains its vehicles. We construct the predicted maintenance cost for a household based on its household characteristics and geographical location. See the appendix for a detailed discussion of the construction of this predicted variable.

4 Basic Model

For clarity, we first introduce a simplified model that abstracts away from sample-selection problems and how private information affects maintenance decisions. A household owns a set of vehicles at any given time, and its composition changes over time. Let d_{itmjs} be a dummy for whether the mth vehicle ever owned by household i during the sample period is owned in quarter t and is a brand j vehicle of age s. The household's cost of maintaining this vehicle depends on its observed and unobserved quality. Thus, we model the cost of maintaining a vehicle as a function of its brand, age, interaction between brand and age, and an error term capturing the effects of unobserved characteristics of the vehicle.[8] The error is the sum of four components: 1) a household-specific error, 2) a household-time-specific error,[9] 3) a household-vehicle-specific error, and 4) an idiosyncratic error. The maintenance cost function is[10]

$$\begin{aligned} c^*_{itmjs} &= \gamma^b_{1j} + \gamma^a_{1s} + \gamma^{ab}_{1js} + \varepsilon_{1itmjs}; \\ \varepsilon_{1itmjs} &= \eta_{1i} + \delta_{1it} + \xi_{1im} + \nu_{1itmjs}; \\ \eta_{1i} &\sim iid\left(0, \sigma^2_{\eta 1}\right); \\ \delta_{1it} &\sim iid\left(0, \sigma^2_{\delta 1}\right); \\ \xi_{1im} &\sim iid\left(0, \sigma^2_{\xi 1}\right); \\ \nu_{1itmjs} &\sim iid\left(0, \sigma^2_{\nu 1}\right). \end{aligned} \tag{1}$$

By aggregating over all vehicles owned by the household, we can define

$$n^b_{itj} = \sum_{m,s} d_{itmjs},$$

$$n^a_{its} = \sum_{m,j} d_{itmjs}, \text{ and}$$

$$n^{ab}_{itjs} = \sum_{m} d_{itmjs}$$

[8] We do not specify costs as a function of model year because we cannot separately identify age, model-year, and time effects.

[9] In the empirical section, δ_{1it} is split into two time effects: a random effect associated with the season and another random effect associated with the calendar year.

[10] The subscript numeral 1 is used to represent maintenance costs parameters, while the subscript numeral 2, introduced later in this section, represents selling decision parameters that are introduced later in the model.

as the number of vehicles of brand j owned by household i in time t, the number of vehicles of age s owned by household i in time t, and the number of vehicles of brand j and age s owned by household i in time t, respectively. Because we include log income as a household characteristic, the cost specification is consistent with a PIGLOG budget shares where brand dummies and region and urban influence annual miles driven, i.e. driving intensity, and thus maintenance decisions.

We aggregate across household vehicles, adjust for household characteristics, and define total household maintenance expenditures, c_{it}, to be

$$\begin{aligned}
c_{it} &= \gamma_{1t} + w_{1it}\beta_1 + \sum_{m,j,s} d_{itmjs} c^*_{itmjs} \qquad (2) \\
&= \gamma_{1t} + w_{1it}\beta_1 + \sum_{m,j,s} d_{itmjs} \left(\gamma^b_{1j} + \gamma^a_{1s} + \gamma^{ab}_{1js} + \varepsilon_{1itmjs}\right) \\
&= \gamma_{1t} + w_{1it}\beta_1 + \sum_{j} \gamma^b_{1j} n^b_{itj} + \sum_{s} \gamma^a_{1s} n^a_{its} + \sum_{j,s} \gamma^{ab}_{1js} n^{ab}_{itjs} + e_{1it}; \\
e_{1it} &= \sum_{m,j,s} d_{itmjs} \left(\eta_{1i} + \delta_{1it} + \xi_{1im} + \nu_{1itmjs}\right)
\end{aligned}$$

where w_{1it} is a vector of household characteristics.[11] To reduce the number of parameters to estimate, the γ^a_{1s} and γ^{ab}_{1js} terms are estimated as splines (with nodes at age 5 and 10).[12] We set the first node at age 5, albeit arbitrarily, to reflect that vehicles require minimal routine maintenance, such as oil changes, in the first 5 years. The second node at age 10 captures the fact that vehicles require more expensive maintenance that is sporadic in nature with more items potentially failing and needing to be replaced, particularly in their final years.

Finally, the covariance matrix terms for e_{1i} implied by the error structure in

[11] We choose not to include interactions between household and car characteristics because doing so require too many degrees of freedom. We choose not to model log(cost) because the data requires aggregation over cars within a household.

[12] The spline specification, however, prevents us from identifying the precise age at which the derivative changes sign.

equation (1) are

$$
\begin{aligned}
Var\left(e_{1it}\right) &= E\left[\sum_{m,j,s} d_{itmjs}\left(\eta_{1i}+\delta_{1it}+\xi_{1im}+\nu_{1itmjs}\right)\right]^2 \qquad (3)\\
&= n_{it}^2\left(\sigma_{\eta_1}^2+\sigma_{\delta_1}^2\right)+n_{it}\left(\sigma_{\xi_1}^2+\sigma_{\nu_1}^2\right)\\
Cov\left(e_{1it},e_{1i\tau}\right) &= E\left[\sum_{m,j,s} d_{itmjs}\left(\eta_{1i}+\delta_{1it}+\xi_{1im}+\nu_{1itmjs}\right)\right]\cdot\\
&\quad\left[\sum_{m,j,s} d_{i\tau mjs}\left(\eta_{1i}+\delta_{1i\tau}+\xi_{1im}+\nu_{1i\tau mjs}\right)\right]\\
&= n_{it}n_{i\tau}\sigma_{\eta_1}^2+\sum_m d_{itm}d_{i\tau m}\sigma_{\xi_1}^2 \quad \text{for } t\neq\tau
\end{aligned}
$$

where

$$n_{it}=\sum_m d_{itm}$$

and $d_{itm}=\sum_{j,s} d_{itmjs}$ which is unity iff vehicle m is owned by i in period t. We assume the error terms in equation (1) are distributed normally so that we can rely on the fact that sums of normal random variables are also normal.

The costs of maintaining vehicles also influence the household's vehicle ownership decisions. First, we construct and analyze a simple model to develop intuition about the household decision, and then we introduce asymmetric information into the model. For each of the vehicles the household owns, the household must decide each period whether to continue owning it or to sell it. We denote household i's decision to sell[13] vehicle m at time t by $z_{itm}=1$ (i.e., sell) iff

$$
\begin{aligned}
c_{itmjs} &> \tau_{imjs}=\gamma_{2j}^b+\gamma_{2s}^a+\gamma_{2js}^{ab}+w_{2it}\beta_2+\eta_{2i}+\nu_{2itmjs}; \qquad (4)\\
\eta_{2i} &\sim iidN\left(0,\sigma_{\eta_2}^2\right);\\
\nu_{2itm} &\sim iidN\left(0,\sigma_{\nu_2}^2\right).
\end{aligned}
$$

This specification reflects the notion that the threshold for selling a vehicle τ_{imjs} depends on 1) the vehicle's brand, age, and the interaction of these terms; 2) the household demographics (such as living in an urban area and the number of drivers per vehicle) to capture the magnitude of the household's driving needs; and 3) error terms capturing unobserved household selling preferences.

Given equations (2), (3), and (4), the joint probability of the vector of de-

[13] Recall the subscript numeral two represents the selling decision parameters as opposed to the cost parameters in equation (1).

pendent variables for household i is

$$\begin{aligned}
\Pr\left[z_i, c_i\right] &= \int \cdots \int \prod_t \Pr\left[z_{it}, c_{it} \mid \eta_i, \delta_{1it}, \xi_{1im}\right] \cdot \\
&\quad \Pr\left(\eta_i, \delta_{1i}, \xi_{1i}\right) d\delta_{1i} d\xi_{1i} d\eta_{1i} d\eta_{2i}; \\
\Pr\left[z_{it}, c_{it} \mid \eta_i, \delta_{1it}, \xi_{1im}\right] &= \frac{1}{\sqrt{n_{it}}\sigma_{\nu_1}} \phi\left(x_{it}\right) \cdot \\
&\quad \prod_{\{d_{itm}=1\}} \Phi\left(y_{itm}\right) \prod_{\{d_{it-1m}=1, d_{itm}=0\}} \left[1 - \Phi\left(y_{itm}\right)\right]
\end{aligned} \tag{5}$$

where

$$\begin{aligned}
x_{it} &= \left(\sqrt{n_{it}}\sigma_{\nu_1}\right)^{-1} \left[c_{it} - \gamma_{1t} - w_{1it}\beta_1 - \sum_j \gamma_{1j}^b n_{itj}^b - \sum_s \gamma_{1s}^a n_{its}^a \right. \\
&\quad \left. - \sum_{j,s} \gamma_{1js}^{ab} n_{itjs}^{ab} - n_{it}\left(\eta_{1i} + \delta_{1it}\right) - \sum_m d_{itm}\xi_{1im}\right]; \\
y_{itm} &= \left(\sigma_{\nu_1}^2 + \sigma_{\varsigma_1}^2\right)^{-1/2} \left(\gamma_{1j}^b + \gamma_{1s}^a + \gamma_{1js}^{ab} + \eta_{1i} + \delta_{1it} + \xi_{1im}\right. \\
&\quad \left. - \gamma_{2j}^b - \gamma_{2s}^a - \gamma_{2js}^{ab} - \eta_{2i}\right)\left(2z_{itm} - 1\right).
\end{aligned} \tag{6}$$

Equation (5) would be the likelihood contribution for observation i except that a) there is a selection problem to handle, and b) we want to amend the model to incorporate private information. These two modifications are discussed in the next two sections.

5 Selection Bias

Recall that we must exclude households from our sample who own vehicles manufactured prior to 1986 because the CEX does not provide the manufacturing year for these vehicles. Without the manufacturing year, we cannot establish the link between maintenance costs and vehicle age for these vehicles. As a result, we face a significant selection-bias problem associated with excluding these households from the sample because the decision to own *newer* models may be correlated with maintenance expenditures (e.g., EHSa). The closer a household is sampled relative to 1986, the more likely it is to be excluded and the larger the potential selection bias. Figure 4 shows the extent to which this is true disaggregated by interview year. The two sets of bars represent the frequency of vehicles by calendar year, one for the full CEX sample and the other for our sample. As one can see, our sample is representative of the frequencies in the full sample in the later years but not in the early years of the sample because of our exclusion rule.

The age density for our restricted sample, disaggregated by calendar year, is displayed in Figure 5. Because households that own vehicles manufactured prior

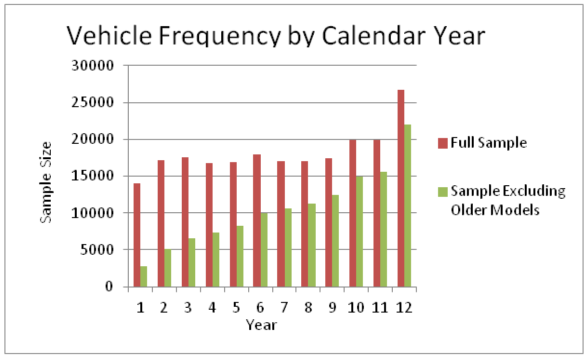

Figure 4: Vehicle Frequency by Calendar Year

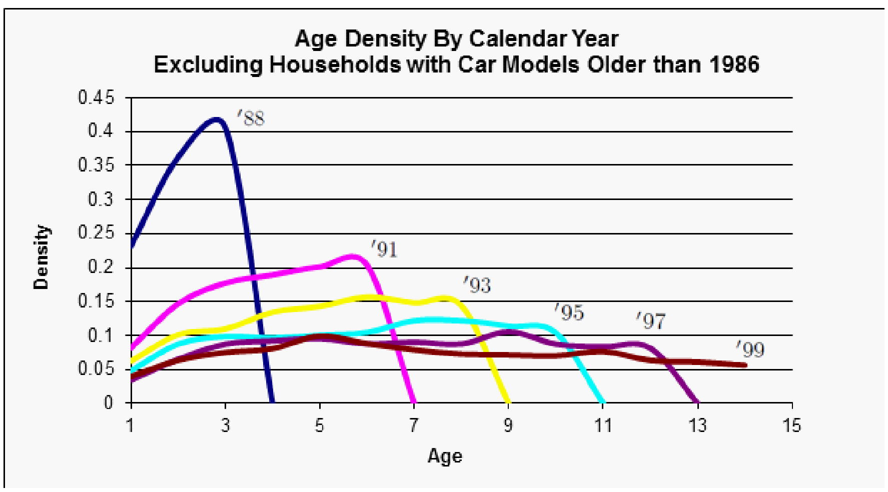

Figure 5: Age Density by Year

to 1986 are excluded from the sample, the oldest vehicle in the 1988 restricted sample is 3 years old, while it is 9 years old for the 1995 restricted sample. It is quite clear that the density curves are moving to higher ages with increasing calendar years.[14]

As in much of the literature, we propose to control for selection bias by modeling selection. Assume that the age of the oldest vehicle in household i in

[14]Computing the density is a bit complicated because manufacture year is aggregated to some degree in years prior to 1986 in the CEX. We estimated a smoothed exponential spline density to handle this problem. Nodes were defined by the aggregation design in the CEX.

the initial period of observation is[15]

$$\begin{aligned} p_i^* &= w_{3i}\alpha + u_i, \\ u_i &= \varphi_i + \zeta_i \end{aligned} \tag{7}$$

with $\zeta_i \sim iidN\left(0, \sigma_\zeta^2\right)$ and φ_i distributed joint normally with η_{1i} in equation (2).[16] The household is included in the sample if $p_i^* \leq \psi_i$ where ψ_i is the maximum age for the oldest vehicle a household could own in that time period and still be manufactured no earlier than 1986. We define a binary selection variable, $p_i = 1\left(p_i^* \leq \psi_i\right)$. In the data, we observe p_i^*, and $\psi(i)$ is known.

There are three standard ways to control for selection bias. The most straightforward is to include the selection bias model directly in the likelihood function (Heckman, 1974; Bloemen, 2008). The problems with this approach are that sometimes the likelihood function is not well-behaved, and there is concern that the estimates are very sensitive to the joint-normality assumption usually made when p_i^* is latent. The second approach is a two-step method (Heckman, 1979). The problems here are that there is still evidence of sensitivity to the joint-normality assumption (Arabmazar and Schmidt, 1982; Olsen, 1982; Stern, 1996), and the method relies on a second step method-of-moments approach that does not fit into our overall estimation methodology. The last approach is semiparametric (Newey, Powell, and Walker, 1990; Ichimura and Lee, 1991; Ichimura, 1993; Stern, 1996). While the semiparametric approach solves the joint-normality sensitivity problem, it does not fit well into our framework for our cost equation.

For our application, the appropriate approach is to incorporate the selection equation directly into the likelihood function. The advantages of this approach are: 1) it is the only approach that is consistent with MLE, and 2) we observe p_i^* as well as p_i thus relying less on the joint-normality assumption.

In much of the literature on controlling for selection bias, there is an implicit belief that one should not rely on a joint-normality assumption to attain identification and instead should use instruments. In fact, work such as Ichimura (1993) shows that, without zero restrictions on some parameters, selection models are not identified nonparametrically. In our case, we do not need instruments because we observe p_i^* as well as p_i; i.e, instead of relying on the functional form of the inverse Mills ratio for identification, we are only estimating an identified correlation between errors from two observed continuous dependent variables. However, instruments may still be useful, providing a source of exogenous vari-

[15] A better model of age of oldest vehicle would look more like a multiple-spell survival model. But that would depend upon the very data elements that we do not observe, for example, the portfolio of vehicles owned at each time in the sample. Our approach is more in the spirit of Heckman and Singer (1984) in that we are specifying a "reduced form" specification for an initial condition.

[16] Thus the error specification in equation (1) changes to

$$\begin{pmatrix} \eta_{1i} \\ \varphi_i \end{pmatrix} \sim iidN\left[0, \begin{pmatrix} \sigma_{\eta_1}^2 & \sigma_{\eta_1\varphi} \\ \sigma_{\eta_1\varphi} & \sigma_\varphi^2 \end{pmatrix}\right].$$

ation. We propose to use predicted maintenance expenditure in the region and predicted age of the oldest household vehicle conditional on other household variables and geography as instruments in the cost and selection equations respectively. Both of these variables can be constructed using just the CEX data.[17] The new likelihood contribution with a selection term is

$$\begin{aligned}\Pr[z_i, c_i] &= \int \cdots \int \Pr[p_i^* \mid \varphi_i] \prod_t \Pr[z_{it}, c_{it} \mid \eta_i, \delta_{1it}, \xi_{1im}]^{p_i} \cdot \\ &\quad \Pr(\eta_i, \delta_{1i}, \xi_{1i}, \varphi_i)\, d\delta_{1i} d\xi_{1i} d\eta_{1i} d\eta_{2i} d\varphi_i\end{aligned}$$

where

$$\Pr[p_i^* \mid \varphi_i] = \frac{1}{\sigma_\zeta}\phi\left(\frac{p_i^* - w_i\alpha_p - \varphi_i}{\sigma_\zeta}\right)$$

and all other terms are already defined. Note that the only necessary information for the excluded sample is $\{p_i^*, w_i\}$ for each household i in the excluded sample.

6 Adding Private Information to the Model

Before we can add private information to the model, we must first clarify what we mean by private information in this context. We are concerned about the availability of information to three parties: the seller/owner, the entire market (i.e. buyers as well as the seller), and the econometrician. We will suppose that the seller can observe all relevant data at least as well as the other two parties. Hence anything relevant that is unobserved by the seller is also unobserved by everyone else, and we can ignore such data. This leaves four categories of information that is observed by the seller, depending on whether it is observed or unobserved by the market (M) or by the econometrician (E):

1. Information observed by the market and by the econometrician, indicated by the letters OM,OE (for example, data like brand model and year);

2. Information observed neither by the market nor by the econometrician- UM,UE (for example, car quality in the classic Akerlof lemons model would fall into this category);

3. Information observed by the market but not by the econometrician- OM,UE (for example, visible dents and scratches, and depending on the richness and quality of the available data sources, data like odometer readings and sales prices);

4. Information observed by the econometrician but not by the market- UM,OE (for example, the cost data that we use in this paper, or at least the component in the variation of costs that is orthogonal to the information that the market does observe).

[17] See the appendix for variable construction.

When we make reference to private information, we are referring to the second and fourth categories, information that is not observed by the market but may be known by the econometrician.

Now we can proceed to add private information to our model. The cost of maintaining a particular vehicle depends on both observed and unobserved characteristics. To the degree that part of a household's maintenance expenditures is not observed by the market, it may influence the timing of the household selling its vehicles. Recall from equation (4) that the household sells a vehicle when the cost of maintaining it exceeds a certain threshold. In order to examine the role information asymmetries play in selling decisions, we decompose the maintenance costs into observed and unobserved components. If the unobserved component drives the selling decisions, this would suggest that asymmetric information is important in the household selling decision.

To determine whether this is the case, we modify equation (4),

$$c_{itmjs} > \tau_{itmjs} = \gamma_{2j}^{b} + \gamma_{2s}^{a} + \gamma_{2js}^{ab} + w_{2it}\beta_2 + \eta_{2i} + \nu_{2itmjs},$$

to allow for asymmetric information with respect to maintenance expenditures. In particular, we redefine the sell decision for household i to become: sell iff

$$\rho c_{itmjs} > \tau_{itmjs} = \gamma_{2j}^{b} + \gamma_{2s}^{a} + \gamma_{2js}^{ab} + w_{2it}\beta_2 + \eta_{2i} + \delta_{2it} + \xi_{2im} + \nu_{2itmjs}$$

where ρ has two components. The first component indicates the degree to which costs, observed and unobserved, influence selling decisions. More specifically $\rho = 0$ implies that cost has no effect on sales propensity. This would occur if all variation in cost were priced in the market and cost had no extra effect on utility. The second component of ρ can be interpreted as translating cost units (measured in \$1000) into standard deviations of the error term in the sell equation, a common feature in choice models. One, however, cannot separately identify these two components of ρ. Even though the first component must be between 0 and 1, the second component only must be positive. We definitively can say only that, if $\rho = 0$, then cost has no effect on sales propensity.

Also note that, in our modified equation above, we add the unobservable heterogeneity terms $\delta_{2it} + \xi_{2im}$ to the sale side of the inequality, terms parallel to $\delta_{1it} + \xi_{1im}$ in the household maintenance cost side of the inequality, so that we are not relying on their absence to identify ρ.[18] In addition, we allow for possible correlation between the error terms across the maintenance and selling decisions [i.e., (η_{1i}, η_{2i}), $(\delta_{1it}, \delta_{2it})$, and (ξ_{1im}, ξ_{2im})].

We further rewrite the selling condition by: a) substituting equation (1) for c_{itmjs} (i.e., the cost of maintaining a vehicle excluding the effect of household characteristics on household aggregate costs) and b) substituting $\eta_{1i} + \delta_{1it} +$

[18] Without the addition of these terms, the standard deviations of δ_{1it} and ξ_{1im} are identified by correlation of cost residuals, and ρ is then identified by the degree to which sell residuals have correlation patterns consistent with those that identified the standard deviations of δ_{1it} and ξ_{1im}. The inclusion of δ_{2it} and ξ_{2im} allows those patterns to be explained by the standard deviations of δ_{2it} and ξ_{2im} rather than ρ.

$\xi_{1im} + \nu_{1itmjs}$ for ε_{1itmjs}. With these modifications, the selling decision becomes[19] [20]

$$\begin{aligned} & \rho\left(\gamma_{1j}^{b} + \gamma_{1s}^{a} + \gamma_{1js}^{ab} + \eta_{1i} + \delta_{1it} + \xi_{1im} + \nu_{1itmjs}\right) \\ > \quad & \gamma_{2j}^{b} + \gamma_{2s}^{a} + \gamma_{2js}^{ab} + w_{2it}\beta_{2} + \eta_{2i} + \delta_{2it} + \xi_{2im} + \nu_{2itmjs}. \end{aligned} \tag{8}$$

Define

$$\begin{aligned} z_{itmjs} \quad = \quad & \rho\left(\gamma_{1j}^{b} + \gamma_{1s}^{a} + \gamma_{1js}^{ab} + \eta_{1i} + \delta_{1it} + \xi_{1im}\right) \\ & -\gamma_{2j}^{b} + \gamma_{2s}^{a} + \gamma_{2js}^{ab} + w_{2it}\beta_{2} + \eta_{2i} + \delta_{2it} + \xi_{2im}), \end{aligned}$$

so that the probability of observing a sale, conditional on cost, is

$$\Pr\left[\nu_{2itmjs} < z_{itmjs} \mid \nu_{1itmjs}\right] = \Phi\left[\frac{z_{itmjs} - \mu\left(\nu_{1itmjs}\right)}{\sigma^{*}}\right]$$

where

$$\mu\left(\nu_{1itmjs}\right) \quad = \left(\frac{\sigma_{\nu 1 \nu 2}}{\sigma_{\nu 1}^{2}} - \rho\right)\nu_{1itmjs}$$

is the mean of $(\nu_2 - \rho\nu_1) \mid \nu_1$, and

$$\sigma^{*} = \sigma_{\nu 2}\sqrt{1 - \left(\frac{\sigma_{\nu 1 \nu 2}}{\sigma_{\nu 1}}\right)^{2}} \tag{9}$$

[19] One might expect the cost of maintaining a vehicle, c_{itmjs}, to be a function of household demographics. Given our empirical specification, we can identify only the difference in how household demographics affect the selling threshold versus the cost of maintaining a vehicle multiplied by ρ.

[20] Equation (8) models the sales decision as determined by the comparison between a multiple of the total maintenance cost and a selling threshold. Instead of using a multiple of the total maintenance costs on the left-hand side, one might decompose this into the part of maintenance costs that is accounted for by observables and the residual part. We might then replace the multiple of total maintenance costs on the left by a linear combination of the two parts of this decomposition, namely the part of maintenance costs accounted for by observables and the residual. Because we allow for correlation in the error terms on the left and right hand side of (8), it turns out that this alternative specification can be rewritten to show that it is essentially the same as the original one.

The new specification has the owner selling iff

$$\begin{aligned} & \rho_{o}\left(\gamma_{1j}^{b} + \gamma_{1s}^{a} + \gamma_{1js}^{ab} + \eta_{1i} + \delta_{1it} + \xi_{1im} + \nu_{1itmjs2}\right) + \rho_{u}\nu_{1itmjs1} \\ > \quad & \gamma_{2ijs} = \gamma_{2j}^{b} + \gamma_{2s}^{a} + \gamma_{2js}^{ab} + w_{2it}\beta_{2} + \eta_{2i} + \delta_{2it} + \xi_{2im} + \nu_{2itmjs}. \end{aligned}$$

For example, if cost components observed by the market have no effect on sales but unobserved components do, then $\rho_o = 0$ and $\rho_u = 1$. However, we can rewrite the inequality as

$$\begin{aligned} & \rho_{o}\left(\gamma_{1j}^{b} + \gamma_{1s}^{a} + \gamma_{1js}^{ab} + \eta_{1i} + \delta_{1it} + \xi_{1im} + \nu_{1itmjs1} + \nu_{1itmjs2}\right) \\ > \quad & \gamma_{2ijs} = \gamma_{2j}^{b} + \gamma_{2s}^{a} + \gamma_{2js}^{ab} + w_{2it}\beta_{2} + \eta_{2i} + \delta_{2it} + \xi_{2im} + \nu_{2itmjs} - (\rho_{u} - \rho_{o})\,\nu_{1itmjs1} \end{aligned}$$

and then just redefine the error associated with sales as $\xi_{2im} - (\rho_u - \rho_o)\,\nu_{1itmjs1}$ to get the same model.

is the standard deviation of $\nu_2 \mid \nu_1$. Because of the usual binary discrete choice variance identification problem, without loss of generality, we set $\sigma^* = 1$, and our probability becomes

$$\Phi\left[z_{itmjs} - \mu\left(\nu_{1itmjs}\right)\right]. \tag{10}$$

The introduction of private information to the model also has implications for the error terms. First, we decompose the error in the cost equation into two components:[21]

- $\nu_{1A} \sim iidN\left(0, \sigma^2_{\nu_{1A}}\right)$, which is observed in the market but not by the econometrician (OM,UE); and
- $\nu_{1B} \sim iidN\left(0, \sigma^2_{\nu_{1B}}\right)$, which is observed only by the car owner (UM,UE).

We do not allow for correlation between ν_{1A} and ν_{1B} because a nonzero correlation would be inconsistent with the definition of ν_{1B} and it is not identified by our data. Recall that the first component of ρ defines how much cost affects sell decisions and therefore should be included in the sell equation on the cost coefficients. Given that, we also decompose the error in the sell equation into three components:

- $\nu_{2A} = \rho\nu_{1A}$, which is OM,UE and captures heterogeneity in preferences over cars with different observable maintenance profiles;
- $\nu_{2B} = \rho\nu_{1B}$, which is the effect of heterogeneous preferences on car characteristics affecting the private component of maintenance profiles and observed only by the owner (UM,UE); and
- $\nu_{2C} \sim iidN\left(0, \sigma^2_{\nu_{2C}}\right)$, which is the effect of heterogeneous preferences associated with selling independent of car characteristics that is observed only by the owner.

Then, the variance and covariance terms are

$$\begin{aligned}
\sigma^2_{\nu_1} &= Var\left(\nu_1\right) = Var\left(\nu_{1A} + \nu_{1B}\right) = \sigma^2_{\nu_{1A}} + \sigma^2_{\nu_{1B}}; \\
\sigma^2_{\nu_2} &= Var\left(\nu_2\right) = Var\left(\nu_{2A} + \nu_{2B} + \nu_{2C}\right) = \sigma^2_{\nu_{2C}} + \rho^2\left(\sigma^2_{\nu_{1A}} + \sigma^2_{v_{1B}}\right); \\
\sigma_{\nu_1,\nu_2} &= Cov\left(\nu_1, \nu_2\right) = Cov\left(\nu_{1A} + \nu_{1B}, \nu_{2A} + \nu_{2B} + \nu_{2C}\right) \\
&= Cov\left(\nu_{1A} + \nu_{1B}, \nu_{2C} + \rho\nu_{1A} + \rho\nu_{1B}\right) = \sigma_{\nu_{1B},\nu_{2c}} + \rho\left(\sigma^2_{\nu_{1A}} + \sigma^2_{\nu_{1B}}\right).
\end{aligned} \tag{11}$$

The question remains how to quantify the effect of private information in the model. Note that ν_{1B} is the random component of cost observed only by the

[21] For a similar decomposition of the error term in auction models, see Athey and Haile (2002), Campo, Perrigne, and Vuong (2003) and Krasnokutskaya (2012). Likewise, for a nonparametric identification of games with unobservable heterogeneity and private information, see Aradillas-Lopez (2010), Bajari, et al. (2010), de Paula and Tang (2011), and Khan and Nekipelov (2012). See Aguirregabiria and Mira (2007) for a dynamic discrete game of incomplete information with both unobservable heterogeneity and private information.

owner, and $\nu_2 = \nu_{2A} + \nu_{2B} + \nu_{2C}$ is the random component in the sell decision. We can think of the effect of a private cost shock as $E\nu_2 \mid \nu_{1B}$. Note that

$$\begin{aligned} E\nu_2 \quad \mid \quad \nu_{1B} &= \frac{\rho\sigma^2_{\nu_{1B}} + \sigma_{\nu_{1B},\nu_{2C}}}{\sigma^2_{\nu_{1B}}}\nu_{1B} \\ &= \left(\rho + \frac{\sigma_{\nu_{2C}}}{\sigma_{\nu_{1B}}} corr\left(\nu_{1B}, \nu_{2C}\right)\right)\nu_{1B}. \end{aligned}$$

Then a measure of the effect of the private information shocks to cost relative to the total randomness in the sell equation is

$$\begin{aligned} \frac{Std_{\nu_{1B}}\left[E\nu_2 \mid \nu_{1B}\right]}{Std\left(\nu_2\right)} &= \left(\rho + \frac{\sigma_{\nu_{2C}}}{\sigma_{\nu_{1B}}} corr\left(\nu_{1B}, \nu_{2C}\right)\right)\frac{\sigma_{\nu_{1B}}}{\sigma_{\nu_2}} \\ &= \frac{\rho\sigma_{\nu_{1B}} + \sigma_{\nu_{2C}} corr\left(\nu_{1B}, \nu_{2C}\right)}{\sigma_{\nu_2}}. \end{aligned} \tag{12}$$

7 Identification

7.1 Identification of Parameters

Define $\gamma_1 = \left(\gamma^b_{1j}, \gamma^a_{1s}, \gamma^{ab}_{1js}\right)$ and γ_2 analogously. From covariation of cost and covariates and second moments of residuals associated with cost, we can identify $\left(\gamma_1, \sigma^2_{\eta 1}, \sigma^2_{\delta 1}, \sigma^2_{\xi 1}, \sigma_{\nu 1}\right)$. This implies the remaining unindentified parameters are

$$\theta_u = \left(\rho, \gamma_2, \beta_2, \sigma_{\eta 2}, \sigma_{\nu 1A}, \sigma_{\nu 1B}, \sigma_{\nu 1B\nu 2C}, \sigma_{\delta 2}, \sigma_{\xi 2}, \sigma_{\nu 2C}\right).$$

The following four empirical relationships help us identify some of the parameters or function of parameters in θ_u:

- Covariation of brand and age of car and household characteristics with sales frequencies identifies $(\rho\gamma_1 - \gamma_2, \beta_2)$;
- Correlation in sales across time within a household identifies $\left(\rho^2\sigma^2_{\eta 1} + \sigma^2_{\eta 2}\right)$;
- Correlation in sales across cars at a particular time with a household identifes $\left(\rho^2\sigma^2_{\delta 1} + \sigma^2_{\delta 2}\right)$;
- The argument used for identification in Elbers and Ritter (1982) applies here as well and identifies $\left(\rho^2\sigma^2_{\xi 1} + \sigma^2_{\xi 2}\right)$.[22]

The following three empirical relationships associated with the variance and covariance terms in equations (11) identify additional parameters:

- $\widehat{\sigma}^2_{\nu_1}$ identifies the sum of the unobserved cost components, $\sigma^2_{\nu_{1A}} + \sigma^2_{\nu_{1B}}$, but does not identify each of these terms separately.

[22] ξ_2 is car-specific unobserved heterogeneity associated with timing of sale, and its distribution is identified the same way as in any hazard model with unobserved heterogeneity.

- Now the correlation of sales decisions and cost residuals identifies $\mu' = \frac{\partial\mu}{\partial\nu_{1B}} = \left(\frac{\sigma^2_{\nu_1,\nu_2}}{\sigma^2_{\nu_1}} - \rho\right)$ which satisfies

$$\widehat{\mu}' = \frac{\sigma_{\nu_{1B},\nu_{2C}} + \widehat{\rho\sigma^2_{\nu_1}}}{\widehat{\sigma^2_{\nu_1}}} - \rho$$

by substituting the covariance term from equations (11) for $\sigma^2_{\nu_1\nu_2}$. Solving for $\sigma_{\nu_{1B},\nu_{2C}}$ results in

$$\sigma_{\nu_{1B},\nu_{2C}} = \widehat{\mu}'\widehat{\sigma^2_{\nu_1}}. \tag{13}$$

- Given the restriction that $\sigma^* = 1$ and equation (11), equation (9) implies

$$\begin{aligned} 1 &= \sigma^2_{\nu_2}\left[1 - \left(\frac{\sigma_{\nu_1,\nu_2}}{\sigma_{\nu_1}}\right)^2\right] \\ \sigma^2_{\nu_2} &= \left[1 - \left(\frac{\sigma_{\nu_1,\nu_2}}{\sigma_{\nu_1}}\right)^2\right]^{-1} \\ &= \left[1 - \widehat{\sigma^2_{\nu_1}}\left(\widehat{\mu}' + \rho\right)^2\right]^{-1}. \end{aligned} \tag{14}$$

Because we cannot identify 11 parameters with 9 restrictions,[23] consider fixing ρ and identifying other parameters as a function of ρ. Conditioning on ρ, the first bulleted identifying relationship listed above allows us to identify γ_2 as

$$\gamma_2 = \rho\widehat{\gamma_1} - \left(\widehat{\rho\gamma_1 - \gamma_2}\right). \tag{15}$$

Similarly, conditioning on ρ allows us to identify $\sigma_{\eta2}$, $\sigma_{\delta2}$ and $\sigma_{\xi2}$ with the second, third, and fourth identifying relationship, respectively:

$$\begin{aligned} \sigma_{\eta2} &= \sqrt{\left(\widehat{\rho^2\sigma^2_{\eta1} + \sigma^2_{\eta2}}\right) - \rho^2\widehat{\sigma^2_{\eta1}}}; \\ \sigma_{\delta2} &= \sqrt{\left(\widehat{\rho^2\sigma^2_{\delta1} + \sigma^2_{\delta2}}\right) - \rho^2\widehat{\sigma^2_{\delta1}}}; \\ \sigma_{\xi2} &= \sqrt{\left(\widehat{\rho^2\sigma^2_{\xi1} + \sigma^2_{\xi2}}\right) - \rho^2\widehat{\sigma^2_{\xi1}}}. \end{aligned}$$

[23] The nine restrictions used to identify parameters can be found in the 7 previous bulleted items. These restrictions contain some vectors of parameters and, for the purpose of counting restrictions, we treat each vector (e.g., γ_1) as 1 restriction. Given this assumption, the first bullet contains two restrictions - one for $(\rho\gamma_1 - \gamma_2)$ and another for β_2. The next 6 bullet items contain 1 restriction each, while the final bullet contains 2 restrictions.

Table 5: Estimated Household Effects

	Selection Effects		Cost Effects		Sales Effects	
	Estimate	Std Err	Estimate	Std Err	Estimate	Std Err
Urban	0.014	0.026	-0.033 **	0.008	-0.029	0.111
Log(Income)			0.067 **	0.001	1.066 **	0.015
# Drivers			0.195 **	0.005	2.544 **	0.044
# drivers/ car	0.700 **	0.011	-0.143 **	0.006		
Avg Maintenance Cost			-0.227 **	0.010	-8.066 **	0.185

In summary, we can identify only a set of parameters and combination of parameters in θ_u conditional on ρ. But, the only parameters in θ_u that are relevant for identifying private information are those in $[\theta_u]^* = (\sigma_{\nu_{1B}}, \sigma_{\nu_{1B},\nu_{2C}}, \sigma_{\nu_{2C}})$, and they must satisfy equations (13) and (14). Thus, for any value of ρ and $\sigma_{\nu_{1B}}$, we can identify the value of $(\sigma_{\nu_{1B}}, \sigma_{\nu_{1B},\nu_{2C}}, \sigma_{\nu_{2C}})$ that satisfies equations (13) and (14). See, for example, Stern (1989) and Koop and Poirier (1997) for a similar type of analysis.

8 Results

8.1 Selection Effects

The selection rule for including households in our sample creates a biased sample. As a result, we estimate equation (7), a selection effect for the age of the oldest vehicle owned by the household. Results are reported in the last column of Table 5. The age of the oldest household vehicle owned is positively related to living in an urban area and the number of household drivers per car. These households are then the ones most likely to be excluded from our sample. As reported in Table 7, the idiosyncratic error σ_ζ in the "oldest household vehicle" selection equation equals 5.977. We conclude from this that there is extensive unaccounted variation in household vehicle portfolio choice with respect to age of the oldest household vehicle owned.

8.2 Maintenance Costs Results

We create maintenance cost profiles by brand and age based on the coefficient estimates for equation (2). The household demographic effects are listed in column 2 of Table 5. Living in an urban area lowers household expenditures by 0.033, approximately $33. Each additional driver raises the cost of maintaining a vehicle but at a decreasing rate with each successive vehicle the household owns. The cost of maintaining a vehicle depends on driving intensity which is a function of the number of drivers as well as the number of vehicles available to drive. Even after controlling for vehicle characteristics and household demographics, we find that households with a unit increase of log income still spend $67 more on their vehicles. Unexpectedly, total household maintenance cost falls as the predicted average maintenance cost for similar households rises in one's region. This

Table 6: Estimated Brand Coefficients

	Brand Cost Effects				Brand Sales Effects			
Brand	Intercept	Age Slope: 0-5	Age-Slope: 5-10	Age Slope: 10+	Intercept	Age Slope: 0-5	Age-Slope: 5-10	Age Slope: 10+
Buick	0.0124	0.0154 **	-0.0021	-0.0207 **	-1.8297 **	0.0725	0.0616 ***	0.0138
Chevrolet	0.0273 *	0.0159 **	-0.0112 **	-0.0071 **	-1.6086 **	0.0361	0.0590 **	-0.0109
Chrysler	0.0404	0.0212 **	-0.0117 *	-0.0097	-2.1158 **	0.1477	0.0551 *	-0.0223
Dodge	0.0163	0.0160 **	-0.0084 *	-0.0114 **	-1.1868 **	-0.0375	0.0853 **	-0.0337
Ford	0.0605 **	0.0120 **	-0.0055 **	-0.0210 **	-1.6692 **	0.0665 **	0.0352 **	0.0014
Geo	0.1214 **	-0.0055	-0.0077		-1.3667 **	-0.0103	0.0460	
Honda	0.1156 **	0.0078 **	-0.0021	-0.0152 **	-1.5367 **	-0.0006	0.0454 **	0.0584 **
Hyundai	-0.0140	0.0436 **	-0.0325 **	0.0318	-1.4489 **	0.0995	-0.0145	0.0640
Mazda	0.0926 **	0.0255 **	-0.0201 **	-0.0146	-1.4893 **	-0.0025	0.1113 **	0.0193
Mercury	0.0101	-0.0201 **	-0.0078 *	-0.0117 **	-1.7960 **	0.1113	0.0493 *	-0.0497 *
Mitsubishi	0.1207 **	0.0057	-0.0146	0.0289	-1.9042 **	0.1517	-0.0900	-0.0560
Nissan	0.0933 **	0.0083	0.0034	-0.0148 *	-1.6246 **	0.0275	0.0534 **	0.0161
Oldsmobile	0.0657 **	0.0098 *	-0.0037	-0.0148 **	-1.5550 **	0.0448	0.0434 **	-0.0016
Plymouth	0.0399	0.0144 **	-0.0252 **	-0.0044	-1.8026 **	0.0979	0.0367	0.0116
Pontiac	0.0353 *	0.0255 **	-0.0109 **	-0.0223 **	-1.5001 **	0.0170	0.0710 **	-0.0318
Saturn	0.0542	0.0245 *	-0.0325		-1.5332 **	0.0008	0.0333	
Subaru	0.0925 *	0.0026	0.0023	-0.0212 *	-1.6466 **	0.0218	0.0647	0.0553
Toyota	0.0595 **	0.0137 **	-0.0069 **	-0.0146 **	-1.7118 **	0.0210	0.0403 *	0.0802 **
Volkswagen	0.1245 **	0.0158 *	-0.0027	-0.0303 **	-1.9228 **	0.1466	0.0113	-0.0399
Volvo	0.0655 **	0.0375 **	0.0057	-0.0431 **	-0.8196 **	-0.1220	0.0748	-0.0410
American Luxury	0.1183 **	0.0176 **	-0.0011	-0.0221 **	-1.5740 **	0.0217	0.0245	0.0244
European Luxury	0.1749 **	0.0408 **	-0.0076 *	-0.0397 **	-1.7053 **	0.0431	0.0428	-0.0367
Japanese Luxury	0.1721 **	0.0087	-0.0155	-0.0493	-0.8931 *	-0.1474	0.0877	0.2837 *
Other	-0.0270	0.0417	0.0201		-0.1029	-0.3496 **	0.0694	
Truck	0.1330 **	0.0120 **	-0.0055 **	-0.0222 **	-1.6220 **	0.0490 **	0.0150	-0.0114

Table 7: Estimated Covariance Terms

	Selection Terms		Cost Terms		Sales Terms	
Household-Specific Effect			-0.092 **	0.002		
Choleski Terms			-0.083 **	0.013	0.048 **	0.020
Season-Specific Effect			0.001	0.002		
Choleski Terms			-0.006	0.021	-0.039 **	0.021
Year-Specific Effect			0.000	0.002		
Choleski Terms			0.129 **	0.018	-0.181 **	0.017
Car-Specific Effect			0.153 **	0.002		
Choleski Terms			0.016	0.011	-0.002	0.014
Sigma(v)			0.365 **	0.000		
Sigma(zeta)	5.977 **	0.021				
Mean Adjustment					0.117 **	0.011

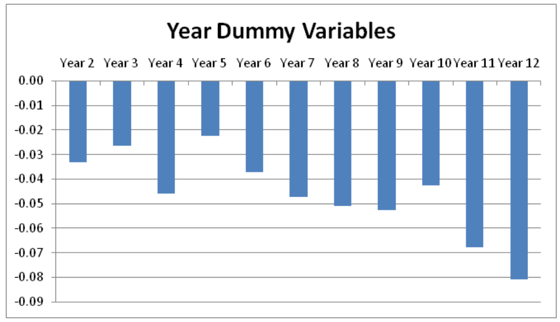

Figure 6: Year Dummy Variables in Maintenance Cost Specification

coefficient may be biased because these predicted costs are a function of regional household demographics but not regional household vehicle composition. For example, this variable does not capture economies of scale present in the auto service and repair business by car type (e.g. domestic versus foreign).

The costs to maintain a vehicle are found to vary over the calendar year and across years. The seasonal dummy estimates suggest that households spend the most on their vehicles during the fall relative to other seasons. The year dummy variables are presented in Figure 6. They reveal that the real cost of maintaining vehicles has fallen over time. This is consistent with aggregate CEX data. The maintenance-related cost components of the Consumer Price Index (CPI) have risen over the sample period, but at a slower rate than the overall CPI (30% *vs.* 41%) (U.S. Statistical Abstract, 1995, 2000).

Table 6 reports the brand-related coefficients in the quarterly maintenance cost equation.[24] Consistent with our priors, the European ($175) and Japanese ($172) luxury brand models are the most costly to maintain, while Buick ($12) and Dodge ($16) are the least costly. Unexpectedly though, the total cost of operating a Honda, on average, is more than a Buick and Dodge. More generally, operating costs of a Honda are higher than for other nonluxury American vehicles. This prediction differs from our initial priors.[25] It is commonly believed that Hondas are of higher quality, and one has to spend less to maintain them. However, the estimated coefficients are consistent with the raw data. In

[24] The test for whether brand effects differ from each other results in a χ^2_{24} test statistic that is 128.42, implying that brand level cost effects are statistically different from each other. This is also true for the brand-age cost slopes; the χ^2_{72} test statistic is 236.67.

[25] *The Complete Car Cost Guide 2000* by Intellichoice tracks seven major ownership cost categories (depreciation, insurance, state fees, financing, fuel, repairs, and maintenance) for the 2000 model-year vehicle. With the exception of the Honda Prelude, the projected 5-year ownership cost for every Honda model is less than for comparable vehicles and often is much cheaper. Conversely, all of the 2000 Buick models have equivalent or higher projected costs than its comparable vehicles.

The J.D. Powers *2003 Initial Quality Survey* provides a measure of vehicle quality for the 2000 model-year vehicle based on the average number of problems experienced within the first 90 days of ownership. In this survey, Honda was the fourth highest-ranking carmaker, while Buick was not even in top 10 and scored below the industry average.

particular, for households that own only one vehicle (and thus all expenses can be attributed to a specific vehicle), households that own a Honda on average spend more on maintenance than households that own a Buick. This might be because the CEX tracks what households spend on their vehicles and not what they should spend on their vehicles to maintain vehicle quality. The unexpected cost of Hondas relative to Buicks may suggest that Hondas have a high resale value (i.e., low depreciation rate) because their owners spend more on maintenance, and thus, when sold, the vehicles are of higher quality.[26]

We also estimate how maintenance costs vary across brands over a vehicle's life. Columns $3-5$ of Table 6 report these brand-age coefficient estimates. They indicate that maintenance cost profiles differ across brands. Except for the Mercury brand, the statistically significant brand-age interactions follow a similar pattern: the total costs of maintaining a vehicle rises with age for the first 5 years of a vehicle's life. Then they fall with a steeper decline after age 10.

Together the estimated brand-related coefficients provide insights on how household expenditures vary by manufacturer, vehicle age, and interaction of these terms. However, a graphical representation of this is more helpful in understanding how costs vary. Because there are 25 vehicle brands, we show only a subset of representative brands. Figure 7 graphically displays the "Age Cost Effects," i.e. how the cost of maintaining a vehicle (in 1000s) changes with age given the coefficient estimates from equation (2). It is a graph of $\widehat{\gamma}^{b}_{1j}+\widehat{\gamma}^{a}_{1s}+\widehat{\gamma}^{ab}_{1js}$ from equation (2) as s varies from 1 to 15 for a particular brand j.

The first cost pattern displayed in Figure 7 is the "Truck" pattern. The costs to maintain trucks rise with age until age 5, level off until age 10, and then fall. This pattern is consistent with our priors. In the first five years, most routine maintenance work is minor, such as tune-ups and oil changes. Then, around the fifth year, depending on driving intensity, vehicles require more expensive, more labor-intensive work such as replacing timing belts and servicing the transmission. Items start to break down and must be repaired or replaced. Pickrell and Schimek (1999) and EHSb suggest that cost declines after age 10 may result from significant reductions in usage of the vehicle.[27] Also, insurance costs decline with age thereby reducing the cost to maintain a vehicle.

[26] Alternatively, one might argue that Hondas are more costly to maintain; imported vehicles have more expensive parts and higher labor costs because there are fewer shops qualified to repair them. Yet, Hondas (and many other non-luxury Japanese vehicles) are different from other import vehicles like BMWs and Volvos. Hondas have a sizeable market presence in the U.S. On average, Hondas alone account for 9% of the car market over the sample period. This is higher than for the total imported luxury market. Furthermore, if one combines Honda, Nissan, and Toyota, the market share is almost a quarter of U.S. sales. Their high popularity has increased the supply of shops to repair them, making their maintenance costs similar to domestic vehicles'.

[27] In this paper, we do not address whether households' decreased usage of older vehicles caused them to spend less maintaining their vehicles or they used their vehicles less because they are more expensive to maintain.

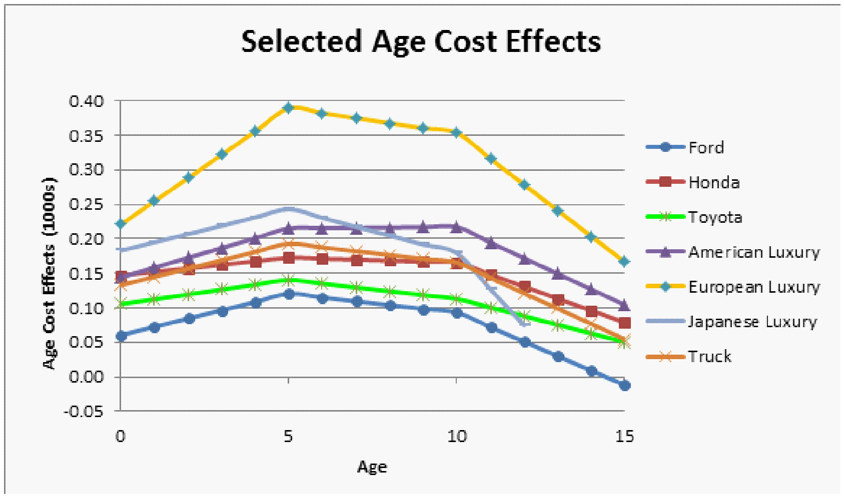

Figure 7: Selected Age Cost Effects

The next patterns displayed in Figure 7 are the "Ford," the "Honda," and the "Luxury" vehicle patterns. The age cost effects are similar to the "Truck" pattern albeit at a lower cost. Toyota and the other American non-luxury brands follow the "Ford" pattern, while Volkswagen follows the "American Luxury" pattern, and Volvo mirrors the "European Luxury" pattern. Note that the cost curves for Japanese luxury vehicles have been truncated. It is not appropriate to extrapolate past age 12 given that the Infinity and Lexus brand did not enter the US market until 1989. Finally with the exception of "Mercury," the remaining brands that have a different shape than presented in Figure 7 are not precisely estimated.

In Table 7, we report the covariance terms for the random effects. The household-specific-cost error σ_{η_1} is 0.092, while the household-vehicle-specific-cost error σ_{ξ_1} is 0.153.[28] This suggests that there is more variation in costs across vehicles within a household than across households. The standard deviations of the household-season-specific-cost error $\sigma_{\delta_{1A}}$ and household-calendar-year-cost error $\sigma_{\delta_{1B}}$ is 0.001 and 0.0002, respective. This implies that there is essentially no variation across households in either time dimension of the maintenance decision. Finally, the standard deviation of the idiosyncratic error is more than twice as large as the car-specific error at 0.365. Thus, while vehicle and household characteristics account for much of the variation in household maintenance decisions, there is still many other factors unobserved to the econometrician that explain household maintenance decisions.

[28] Table 7 reports Choleski terms that can be translated into standard deviations.

8.3 Sales Effects

The second set of columns in Tables 5 and 6 report how various household demographic and vehicle characteristics affect selling decisions. Recall that a household's expenses must exceed a certain threshold in order for it to sell the vehicle. The results indicate a rise in the number of drivers holding total household income constant raises the threshold of selling and therefore lowers the probability of selling a vehicle. One explanation for this finding is that these households are poorer in terms of per person household income. They have insufficient income to sell an existing vehicle to purchase a newer one. Another explanation is that demand for driving services varies across household drivers (e.g., a teenager vs. a working adult). The additional benefit generated from another driver, especially for those nonworking drivers, does not exceed the additional cost of upgrading the household vehicle portfolio by purchasing a newer one. A rise in household income also lowers the probability of selling a vehicle conditional on vehicle brand and age and other demographic characteristics. As with the cost specification, we use log income to allow for diminishing income effects in our sales decisions. Finally, as expected, a household is more likely to sell its vehicle the higher the predicted average cost of maintaining a vehicle in its area.[29]

Unlike in the cost equation, all of the brand effects are not statistically significantly different from each other.[30] This implies that differentiation in product quality across brands is not an important explanatory variable in defining household selling thresholds. Consistent with our priors, the probability of selling a vehicle rises with age. But, vehicle age is not found to be a significant factor in the selling decision after age 10. As with the age cost effects, a graphical representation by brand of how likely a vehicle will be sold over its lifetime is informative.

Figures 8 and 9 demonstrate how the sales threshold changes with age given the coefficient estimates from equation (8).[31] For a particular brand j, we graph the sales threshold conditional on its age which varies from 1 to 15. Note, at age 15, regardless whether the vehicle is a Buick, a Volkswagen, a luxury vehicle, or a truck, there is a convergence in the selling threshold. A similar convergence pattern is observed for the other nonluxury foreign vehicles, albeit at a slightly lower selling threshold.

Table 7 lists the estimated covariance terms between the random variables across the maintenance and selling decisions. The standard deviations of the household-specific error affecting the sales threshold σ_{η_2} equals 0.096.[32] The

---

[29] The sell estimates reported in Tables 5 and 6 are the net effect $(\rho\gamma_1 - \gamma_2)$ from equation (9). Thus, the estimate of γ_2 depends upon the (unidentified) value of ρ. However, given the difference in magnitude between the estimates for γ_1 and γ_2, the estimates of γ_2 are quite insensitive to the true value of ρ. The same is true for the household characteristics estimates.

[30] A test for whether the brand level sales coefficients differ from each other results in a χ^2_{24} test statistic of 30.86. This implies that brand effects are not statistically different from each other. This is also true for the brand-age sales slopes. The χ^2_{72} test statistic is 42.25.

[31] Because there are many brand effects, we arbitrarily report them in two different figures.

[32] Table 7 reports relevant Choleski terms of -0.083 and 0.048. These are combined to get

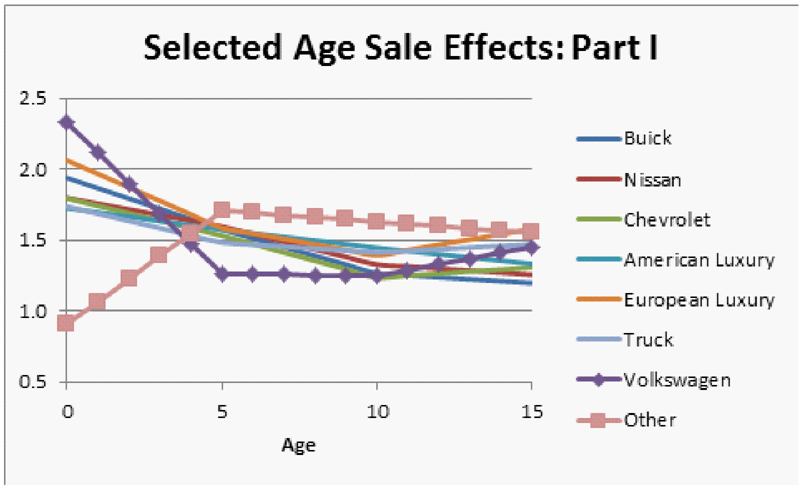

Figure 8: Selected Age Sale Effects Part I

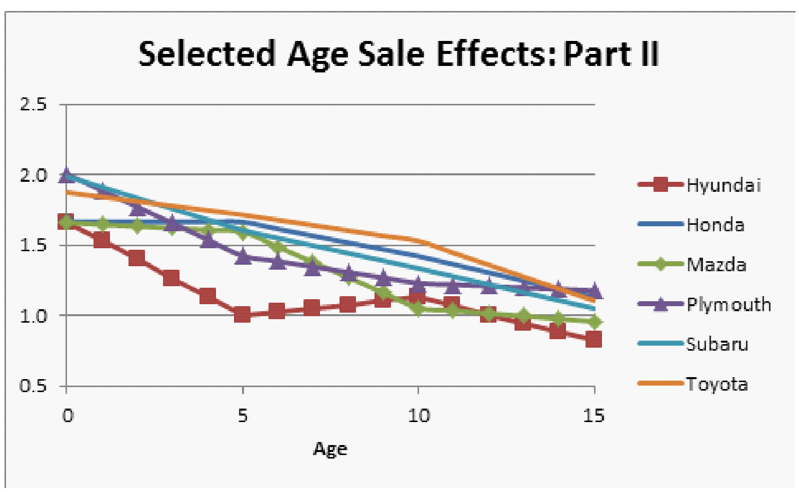

Figure 9: Selected Age Sale Effects Part II

small estimate of the standard deviation suggests that there is small variation in the selling threshold across households due to unobservables. This threshold is found not to vary much across vehicles as well ($\sigma_{\xi_2} = 0.016$). Unlike for the maintenance decision, there is a nonzero time component in the variation in selling decisions across households. The standard deviations of the household-season-specific-sales error $\sigma_{\delta_{2A}}$ and household-calendar-year-sales error $\sigma_{\delta_{2B}}$ are 0.039 and 0.222, respectively. With respect to the covariance between the random variables across the maintenance and selling decisions, the household-specific terms are positively correlated (0.865). The households that spend more on their vehicles have a higher selling threshold. This is inconsistent with EHSa. New car owners who care about car quality are found to maintain their cars and sell them slowly. These results also imply that the sell threshold increases with prior unexpected maintenance costs, suggesting that households who intend to sell spend less on maintenance the period before.[33] However, the size of the standard deviations are small enough to make this a relatively small effect.

The estimates also indicate that there is no correlation between the time-specific random variables across the two decisions. The correlation between the car-specific terms is 0.992 with larger corresponding standard deviations, thus suggesting the joint maintenance/sales decision has a significantly larger car-specific component than household-specific component.

Further analysis of the brand effects in turnover rates can provide additional information on the importance of private information in sales decisions. Hendel and Lizzeri (1999) demonstrate that, under adverse selection, vehicles with greater product variability have lower trade volume and steeper decline in prices. Using data on Kelly Blue Book prices, we estimate the correlation between log price declines and the selling threshold as a vehicle ages. This allows us to determine at what points over the vehicle's life it suffers more from information asymmetry problems. These correlations are shown in Figure 10. One should notice that, for the majority of vehicle ages, the correlations are negative; vehicles with more substantial log price declines are less likely to be sold. These turnover patterns then indicate that the used car market suffers from information asymmetries as applied to Hendel and Lizzeri (1999).[34]

Figure 10 also reveals that private information in the used vehicle market worsens as vehicles age. This result is consistent with the findings by EHSa. However, in this paper, we have more detailed information on how the lemons problem varies over the entire life of the vehicle (and not simply for vehicles older

---

$0.096 = \sqrt{(-0.083)^2 + (0.048)^2}$.

[33] Peterson and Schneider (2011) do not find evidence that owners perform less maintenance in the several quarters prior to vehicle disposal relative to what is performed for similarly continuously held vehicles. Their test for moral hazard, however, cannot distinguish between moral hazard and adverse selection where car problems are due to exogenous reasons causing households to spend more prior to selling. Both effects may cancel each other out and such the authors only can determine that the net effect is not statistically significant.

[34] This is also consistent with Peterson and Schneider (2011). This paper presents a test for adverse selection based on a theoretical model of turnover patterns with respect to defect rates. Vehicles with higher defect rates for engines and transmissions are found to have lower turnover rates, suggesting an asymmetric information problem is present.

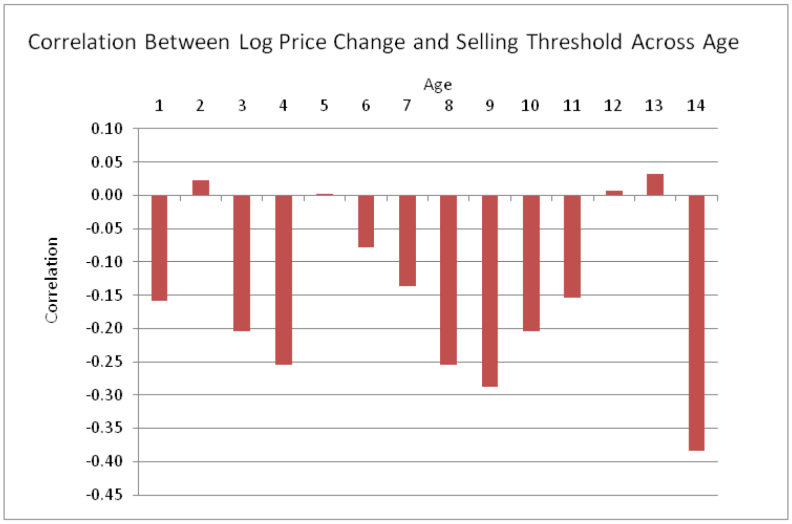

Figure 10: Correlation Between Log Price Change and Selling Threshold Across Age

than 2 years). We see from the figure that the private information effect peaks around age 9 and declines afterwards but rises sharply again at age 14. This is consistent with previous work by Genesove (1993), Bond (1984), and Lacko (1986) who find evidence of private information only among older vehicles.

Using parameter estimates from equation (12), we also can quantify the effect of private information on selling decisions. These effects are shown in Figure 11. Given there are 3 unidentified parameters in equation (12): ρ, $\sigma_{\nu_{1B}}$, and $\sigma_{\nu_{2C}}$, we must consider possible values of ρ and $\sigma_{\nu_{1B}}$ when measuring how much of the unobserved variation in selling is due to private information shocks to cost. The top line demonstrates how $\sigma_{\nu_{2C}}$from equation (12) varies over ρ. Given our estimates of $\widehat{\mu}' = 0.117$ and $\widehat{\sigma}_{\nu_1} = 0.365$ in Table 7, we calculate that $\sigma_{\nu_{2C}}$ explodes at $\rho = 2.623$. Thus, we can limit our analysis to values $0 \leq \rho \leq 2.623$.

The remaining three curves demonstrate the extent to which the market suffers from private information for different values of $\sigma^2_{\nu_{1B}}$. The top convex curve assumes that all unobserved heterogeneity in costs are observed by the market, but not by the econometrician. Thus if $\sigma_{\nu_{1B}} = 0$, then the relative effect of private information is zero. The lowest convex curve assumes that only 1/3 of the unobserved heterogeneity in costs is known only by the owner. Given these alternative values for $\sigma^2_{\nu_{1B}}$, we observe that the share of the total randomness in the sell equation due to private information shocks to cost can be as low as zero and at most of 40%.[35]

[35] We assume that all unobserved problems with the vehicle can be fixed or at least are not extraordinarily hard to fix. Also, there are some unobservable car quality problems that do not affect current maintenance costs, but may later affect the net benefits of owning the car in

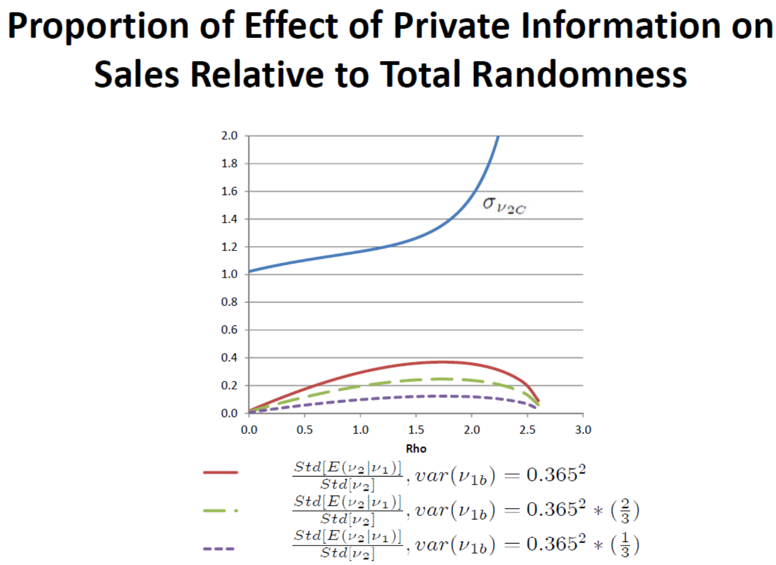

Figure 11: Proportion of Effect of Private Information on Sales Relative to Total Randomness

We also might want to compare our estimated value of private information with other estimates in the literature. Recall that we distinguish private information [(UM,UE) + (UM,OE)] from information observed by the market but not by the econometrician (OM,UE). It is important to decompose unobservable heterogeneity (UH) into these two terms because only the first component is relevant for analyzing how asymmetric information affects markets. For four related studies on asymmetric information, Table 8 lists the market studied, whether the model includes heterogeneity observed by the market but not by the econometrician, and the size of private information. Cardon and Hendel's (2001) model of health insurance does not include heterogeneity observed by the market but not by the econometrician (OM,UE). Because the estimated value of the standard deviation of UH is so small, decomposing UH into two parts would not change their results; the size of private information would still be small. In contrast, EHSa estimates large standard deviations of UH in survivor curves (UH sd = 2.9) for the used car market. It is not clear whether the UH term captures private information and/or information known by the market but not the econometrician. Thus, the extent to which private information drives selling rates depends on the interpretation of the hazard rates and potentially less than previously estimated. Similarly, Finkelstein and Poterba (2004) finds

the future (e.g., an unreported accident that results in structural damage that may adversely affect its drivability and/or maintenance costs in the future.) With such, the unobservable vehicle quality only would influence the sell decision but not the maintenance decision; this component of the error in the sell equation would not be correlated with cost equation error in this instance.

Table 8: Estimates in Literature of Size of Private Information Effects

Paper	Market	Inclusion of heterogeneity observed by market but not by econometrician?	Effects of private Information on Behavior
Cardon & Hendel (2001)	Health Insurance	No	Very Small
EHS (2009a)	Used Cars	Depends‡	Large†
Finkelstein & Poterba (2004)	Annuities	No	Large†
Finkelstein & McGarry (2006)	Long-Term Care Insurance	No*	Small†**

‡ See text for an explanation for what "depends" means.
† Results cannot be quantified to make comparisons across papers.
* Not relevant for model.
** The change in behavior is the sum of 2 separate private information effects that work in opposite directions so the overall effect is negligible. Each individual private effect, however, could be large.

large amount of UH plaguing the annuities market. However, it is unclear what is the true size of private information present in the market because it does not decompose the UH term into private and public information (OM,UE). Finally, Finkelstein and McGarry (2006) estimates a small amount of private information present in the purchases of long-term care insurance. This small amount is the sum of two separate private information effects - private information about the risk of being admitted into a nursing home and private information about preferences for insurance coverage. These two effects work in opposite directions so the overall effect is negligible. Each individual private effect, however, may still be large. The model does not need to account for (OM,UE) given the authors' careful treatment to ensure they utilize the same information in pricing long-term care insurance policies as insurance companies use.

In summary, we have shown that the cost of maintaining a vehicle is one of the avenues in which asymmetric information plagues the used car market. The amount of private information in the used vehicle market worsens as vehicles age, peaking around age 9. Our estimates suggest that the maximum share of the total randomness in the sell equation due to private information shocks to cost is 40% with zero as the lower bound.

9 Correlation of log Cost and log Annual Miles Driven

The cost of maintaining a vehicle depends on how much it is driven. Vehicles driven only 3000 miles per year require less maintenance than those driven 20000 miles per year. Ideally, we would control for driving intensity using annual miles driven. But, in the CEX, we observe miles driven through households' reporting of odemeter readings, and these were found to be unreliable.[36] Instead, we capture how driving intensity affects costs mainly through two variables: number of household drivers and number of drivers per vehicle owned by the household. Driving intensity is also captured by the age variable because households use a vehicle for trips of varying length across its lifetime. For example, households drive their newer vehicles on long road trips, while older vehicles are relegated to short trips in the city (EHSb).

Despite the CEX data limitation, we want to establish the link between annual miles driven and the costs of maintaining a vehicle as it ages. Using data from the 1995 National Personal Transportation Survey (NTPS), we estimate brand-age slopes with respect to log annual miles driven.[37] We combine these with the brand-specific age slopes estimates with respect to cost presented in Table 5. This allows us to analyze the correlation between log maintenance costs and log miles driven annually over a vehicle's life. We have two alternative hypotheses about the relationship: a) Brands with unusually large negative slopes for annual miles driven have large negative slopes for cost. This would imply that households who do not drive their vehicles also do not take care of them. b) Brands with unusually large positive slopes for annual miles driven have large negative slopes for cost. This would suggest that households who do not maintain their vehicles are the ones who drive their vehicles extensively.

In Figure 12, we present brand-specific correlation estimates of log maintenance cost and log annual miles driven over a vehicle's life. For example, the correlation for the Mazda vehicles is 0.9, while it is almost 1.0 for Hondas. Interestingly, the correlation is uniformly positive and large for all of the brands. This implies that, as vehicles age, households that spend large amounts on their vehicles also drive them extensively. This is true across all brands.

We also can examine the relationship between cost and annual miles driven conditional on age. These results are shown in Figure 13. The solid line indicates that the correlation between cost and annual miles driven is 0.5 for new vehicles (i.e., when age equals 0). If one excludes European Luxury from the analysis, the correlation falls to 0.43 as shown by the dashed line. Across brands, the correlation after age 13 continually drops as the vehicle ages. The results suggest households invest in their vehicles for the first 13 years of their lives and drive them extensively. Afterwards, the cars are no longer cost-effective for households to maintain.

10 Alternative Model

The relationship that we find earlier in Section 8.3 between maintenance costs and the likelihood of subsequent sale is consistent with the presence of private

[36] Odometer readings reported in the CEX are unreliable. In fact, 8% of the observations for which one is able to calculate quarterly miles driven using odometer readings imply negative miles driven.

[37] See EHSb for a description of the NTPS.

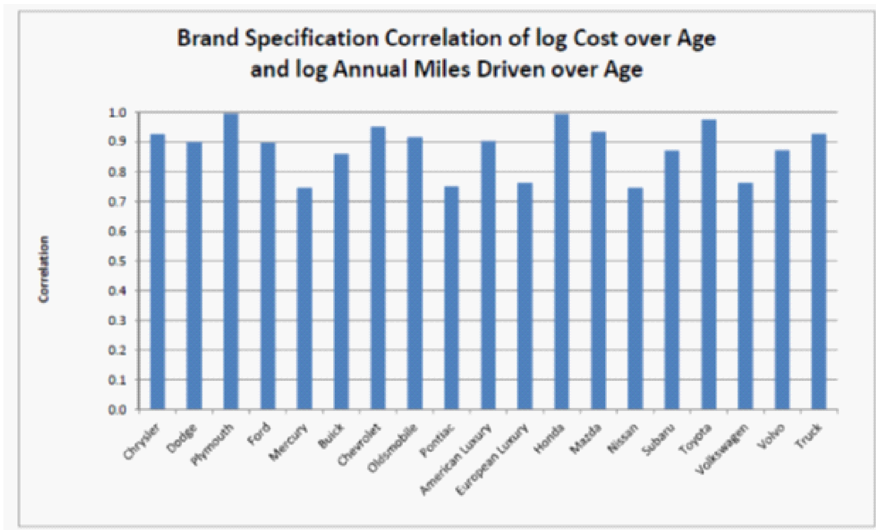

Figure 12: Brand Specification Correlation of log Cost over Age and log Annual Miles Driven over Age

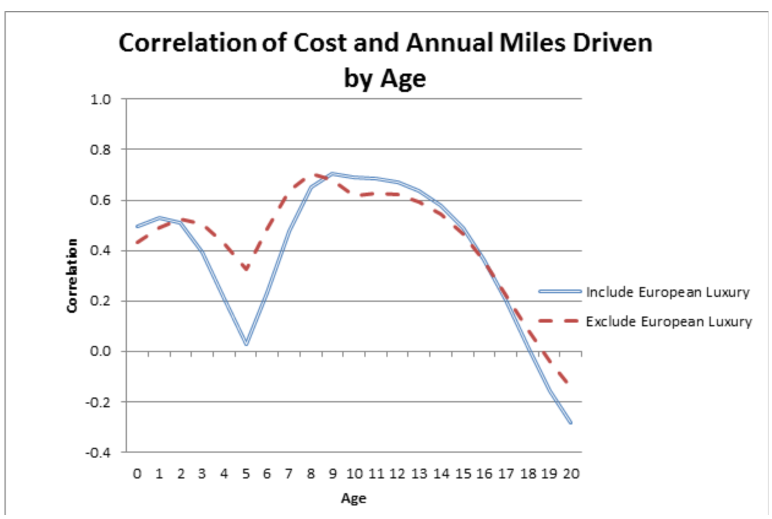

Figure 13: Correlation of Cost and Annual Miles Driven By Age

information about car quality, but there are also possible explanations for this relationship that do that do not rely on private information. One such possibility depends on variations in consumer tastes that could be public information fully observable by everyone in the market, but not by the econometrician.

This problem we have encountered in our paper seems to us to arise quite generally in empirical research on asymmetric information in markets: What might appear to be the effects of asymmetric information between buyers and sellers may instead be due to variables that are fully observed by buyers and sellers but unobserved by the econometrician.[38] Ruling out this latter possibility requires an unusual degree of confidence that everything observed by the parties and used in their decisions is also observed by the econometrician. An example of a paper that makes a serious case that all such information is observed by the econometrician is Finkelstein and McGarry (2006).[39] If one is unsure whether there is some variable that is unobserved by the econometrician but is observed by buyers and sellers, then any relationship in the data that might be interpreted as evidence of asymmetric information can instead be explained by this hidden variable.

The issue can be illustrated by an example. Hendel and Lizzeri (1999) constructs a model of car markets and shows that the presence of asymmetric information has a testable implication: brands of used cars that suffer more from asymmetric information tend to experience steeper price declines and also to have lower volumes of trade. Because in practice one finds a negative correlation across brands between the absolute price decline and the volume of trade, this correlation is taken as evidence of the effects of asymmetric information. But, as we suggested, one can find alternative explanations for this relation if one supposes that buyers and sellers can both fully observe some variable but the econometrician does not.

To keep things simple, suppose there are two periods, two types of cars and two types of consumers. Buyers and sellers can all observe car quality but the econometrician cannot. Suppose that during period 1, brand A cars always decline in quality by a fixed amount (so the variance is zero). Suppose that brand B cars either do not decline in quality at all or, with some positive probability, decline by an amount larger than the decline in quality for brand A cars, so that the mean decline for B is the same as for A. Suppose new-car owners care about quality so that they will sell their car if and only if it declines in quality. Then we see that the volume of trade in brand B cars would be lower than in brand A cars, while the price-drop in the cars traded (which matches the quality decline here because buyers and sellers both observe quality) would be greater for brand B cars too, matching the pattern in the Hendel and Lizzeri

[38] We are grateful to Jose Canals-Cerda for first bringing this possibility to our attention.

[39] Their very careful and convincing work establishes that insurers, in pricing their policies, do not use data such as income which, because it is correlated with private information about individual preferences would help to increase their expected profits. This, however, still leaves open the possibility that such data may nevertheless have some influence on insurer behavior. For example, salespersons may devote more effort selling policies to wealthier clients than to less wealthy ones, even though the terms of the policies sold are the same.

(1999).

Given that our results are consistent with a model without private information, this leads one to question the previously estimated magnitude of private information found in our and other studies in the literature. The inability to identify private information with existing data warrants further attention if we are ever to make policy recommendations with respect to this market failures.

11 Conclusion

In this paper, we have shown that examining the costs that households incur in maintaining their vehicles provides evidence of the importance of asymmetric information in used-car markets. Private information is identified through the link between the unobserved component of household vehicle-maintenance costs (i.e. the part not accounted by household demographics and car characteristics) and the decision to sell. In quantifying private information, we are careful to distinguish information private to the seller (UM, UE) from information known by the market but not by the econometrician (OM,UE). It is important to decompose unobservable heterogeneity into these two terms because only the first component is relevant for analyzing how asymmetric information affects markets and thus, correctly quantifying private information in the market.

We first estimate how household automobile-related expenses vary by vehicles' age, brand, and the interaction of those terms. We observe costs rising in the first 5 years, leveling off until age 10, and then falling afterwards. This pattern is consistent with Pickrell and Schimek (1999) and EHSb. Costs decline after 10 years because older vehicles are driven much less and thus require less costly maintenance. Our findings about how a vehicle's quality endogenously changes over its lifetime promises to be useful for those studying automobile-related issues such as scrapping or buying and selling decisions where households compare the vehicle's sale (scrap) price against the cost of maintaining it.

We then model the household's decision to sell as a function of the vehicle characteristics and the household demographics, both of which affect the household driving intensity and thus need for the vehicle. A household sells the vehicle when the cost of maintenance exceeds a certain threshold. We allow for correlation in the error terms across the household maintenance and selling decision. We find evidence that the unobservable components in the maintenance decision influence selling decisions, suggesting asymmetric information is present. In particular, we observe that households with unusually low maintenance costs (relative to what are expected conditioned on household demographics and vehicle age and brand) are more likely to sell their car. If a household intends to sell a vehicle, these expenses are not worth undertaking when they cannot be capitalized in the price. We also observe vehicles with more substantial log price declines are the ones that are less likely to be sold. This provides additional evidence that private information is present in the market. This information problem is found to worsen as cars age, peaking at age 9.

We have modeled the household's decision to dispose of a vehicle as determined by a static comparison between current maintenance costs and a household-determined threshold level. A more careful dynamic analysis comparing the future costs and benefits of retaining the vehicle remains a goal for future work.

Further attention also warranted for determining how to identify private information with existing data. Our results are consistent with a model without private information. Markets previously found in the literature suffering from private information actually may have catured something else. A more careful analysis of operating cost profiles and how they influence selling decisions when vehicle quality is partially unobservable is necessary to make more accurate policy recommendations.

12 Appendix

We construct two household variables to act as instruments for identification of parameters in the selection model. These two instruments are essentially predicted values of two outcome variables conditional on other household variables and region of residence. The vector of household demographics, w_{it}, can be decomposed into three parts. w_{it1} and w_{it2} are the two constructed variables, predicted household values based on CEX data, while w_{it3} represent the remaining set of household variables that are observed in the CEX. The observed variables include the number of drivers, household size, a dummy for whether the household lives in an MSA, and the household income. In the CEX, we also observe the region of residence.[40] Using data from all households, we run an OLS regression of total household vehicle maintenance cost on all of the variables in w_{it3}, excluding the MSA dummy. A separate regression is done for each region and calendar year. Then using these estimated coefficients, for each household we construct predicted total maintenance expenditures given their household demographics and region of residence. The set of these predicted values form w_{it1}. For the selection problem, we perform a similar procedure where the dependent variable in the OLS regression is age of the oldest household vehicle and then set w_{it2} equal to the predicted value based on the estimates from those regressions. Table 9 and 10 provide information on the descriptive statistics for the data used to construct the average predicted household maintenance and age of the oldest household vehicle, respectfully.

[40] The CEX reports the household's state of residence only after 1992. Also, for some low population states and for some unspecified subsets of other states, we observe only the Census region. Given that we observe Census region in all years, we choose to use this more aggregate geographic variable instead. The four regions of the United States are defined as 1) Northeast, 2) Midwest, 3) South, and 4) West.

Table 9: Moments for the Synthetic Variable Average Maintenance Cost

Variable	# Obs	Mean	Std Dev	Minimum	Maximum
Total					
Average maintenance cost	47206	0.655	0.247	0.209	3.305
By Region: Average maintenance cost					
Northeast	8423	0.660	0.271	0.209	2.322
Northwest	11237	0.608	0.232	0.222	1.912
South	15595	0.638	0.225	0.234	1.895
West	11951	0.716	0.258	0.229	3.305
By Family Characteristics: Average maintenance cost					
Family size < = 2	28131	0.561	0.200	0.209	1.380
Family size > 2	19075	0.793	0.245	0.229	3.305
# of drivers <= 2	39693	0.600	0.208	0.209	1.409
# of drivers > 2	7513	0.943	0.237	0.370	3.305
Log Income <= 6	22566	0.678	0.257	0.209	3.305
Log Income > 6	24640	0.633	0.236	0.219	2.048
Household lives outside of a MSA	5934	0.550	0.209	0.209	1.791
Household lives in a MSA	41272	0.670	0.248	0.245	3.305

Notes: All maintenance costs and income figures are in 1000s

Table 10: Moments for Age of Oldest Vehicle Owned

Variable	# Obs	Mean	Std Dev	Minimum	Maximum
Total					
Average oldest car owned	47206	1.179	0.163	0.819	2.506
Average oldest car owned by region					
Northeast	8423	1.007	0.106	0.819	1.725
Northwest	11237	1.196	0.138	0.898	2.379
South	15595	1.169	0.136	0.920	2.097
West	11951	1.300	0.139	1.038	2.506
Average oldest car owned condition on family characteristics					
Family size < = 2	28131	1.119	0.13	0.819	1.565
Family size > 2	19075	1.268	0.167	0.825	2.506
# of drivers <= 2	39693	1.144	0.134	0.819	1.648
# of drivers > 2	7513	1.365	0.178	0.918	2.506
Log Income <= 6	22566	1.136	0.158	0.819	2.379
Log Income > 6	24640	1.219	0.158	0.894	2.506
Household lives outside of a MSA	5934	1.216	0.171	0.819	2.019
Household lives in a MSA	41272	1.174	0.161	0.842	2.506

Notes: All maintenance costs and income figures are in 1000s

13 References

References

[1] Aguirregabiria, Victor and Pedro Mira (2007). “Sequential Estimation of Dynamic Discrete Games.” *Econometrica*. 75(1): 1–53.

[2] Akerlof, George (1970). “The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism.” *Quarterly Journal of Economics*. 84(3): 488-500.

[3] Alberini, Anna, Winston Harrington, and Virginia McConnell (1996). “Estimating An Emissions Supply Function From Accelerated Vehicle Retirement Programs.” *Review of Economics and Statistics*. 78(2): 251-265.

[4] Alberini, Anna, Winston Harrington, and Virginia McConnell (1998). “Fleet Turnover and Old Car Scrap Policies.” *Resources for the Future Discussion Paper 98-23*.

[5] Arabmazar, Abbas and Peter Schmidt (1982). “An Investigation of the Robustness of the Tobit Estimator to Non-Normality.” *Econometrica*. 50: 1055–1063.

[6] Aradillas-Lopez, A. (2010). “Semiparametric Estimation of a Simultaneous Game with Incomplete Information.” *Journal of Econometrics*. 157: 409-431.

[7] Athey, S. and P. Haile (2002). “Identification of Standard Auction Models.” *Econometrica*. 70(2): 2107-2140.

[8] Bajari, P., H. Hong, M. Krainer, and D. Nekipelov (2010). “Estimating Static Models of Strategic Interactions.” *Journal of Business Economics and Statistics*. 28(4): 469-482.

[9] Bensoussan, Alain, Qi Feng, and Suresh Sethi (2011). “Integrating Equipment Investment Strategy with Maintenance Operations under Uncertain Failures.” Unpublished manuscript.

[10] Bloemen, H. (2008). “Job Search, Hours Restrictions, and Desired Hours of Work.” *Journal of Labor Economics*. 26(1): 137-180.

[11] Blundell, R. and A. Duncan (1998). “Kernel Regression in Empirical Microeconomics.” *Journal of Human Resources*. 33(1): 62-87.

[12] Blundell, R., P. Pashardes, and G. Weber (1993). “What Do We Learn about Consumer Demand Patterns from Micro Data?” *American Economic Review*. 83(3): 570-597.

[13] Börsch-Supan, Axel and Vassilis Hajivassiliou (1993). "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models." *Journal of Econometrics*. 58(3): 347-68.

[14] Bond, Eric (1984). "Test of the Lemons Model: Reply." *American Economic Review*. 74(4): 801-804.

[15] Campo S., Perrigne I., and Q. Vuong (2003). "Asymmetric Bidders in First-Price Auctions with Affiliated Private Values." *Journal of Applied Econometrics*. 18(2):179-207.

[16] Cardon, James and Igal Hendel (2001). "Asymmetric information in Health Insurance: Evidence from the National Medical Expenditure Survey." *Rand Journal of Economics.* 32(3): 408-27.

[17] Chen, Qi (2011). "Optimization Model of Equipment Maintenance." *Mechanic Automation and Control Engineering*. 1463 - 1466.

[18] Cho, Sungjin and John Rust (2008). "Is Econometrics Useful for Private Policy Making? A Case Study of Replacement Policy at an Auto Rental Company." *Journal of Econometrics*. 145(1/2): 243-257.

[19] de Paula, A. and X. Tang (2011). "Inference of Signs of Interaction Effects in Simultaneous Games with Incomplete Information." *Econometrica*. 80(1): 143–172.

[20] Deaton, A. and J. Muellbauer (1980). "An Almost Ideal Demand System." *American Economic Review*. 70(3): 312-326.

[21] Elbers, Chris and Geert Ridder (1982). "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazards Model." *Review of Economic Studies*. 49(3): 403-409.

[22] Emons, Winand and George Sheldon (2009). "The Market for Used Cars: New Evidence of the Lemons Phenomenon." *Applied Economics*, 41(22): 2867-2885.

[23] Engers, Maxim, Monica Hartmann, and Steven Stern (2009a). "Are Lemons Really Hot Potatoes?" *International Journal of Industrial Organization*, 27(2): 250-263.

[24] Engers, Maxim, Monica Hartmann, and Steven Stern (2009b). "Mileage Drives Used Car Prices." *Journal of Applied Econometrics*. 24(1): 1-33.

[25] Esteban, Susanna and Matthew Shum (2007). "Durable Goods Oligopoly with Secondary Markets: the Case of Automobiles." *Rand Journal of Economics*, 38(2): 332-354.

[26] Fabel, Oliver and Erik E. Lehmann (2002). "Adverse Selection and Market Substitution by Electronic Trade." *International Journal of Economics of Business*. 9(2): 175-193.

[27] Finkelstein, Amy and James Poterba (2004). "Adverse Selection in Insurance Markets: Policy Holder Evidence from the UK Annuity Market." *Journal of Political Economy.* 112 (1): 183–208.

[28] Finkelstein, Amy and Kathleen McGarry (2006). "Multiple Dimensions of Private Information: Evidence from the Long Term Care Insurance." *American Economic Review.* 96 (4): 938–958.

[29] Genesove, David (1993). "Adverse Selection in the Wholesale Used Car Market." *Journal of Political Economy*. 101(4): 644-665.

[30] Geweke, John (1988). "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference." *Journal of Econometrics*. 38(1/2): 73-89.

[31] Heckman, James (1974). "Shadow Prices, Market Wages, and Labor Supply." *Econometrica*. 42(4): 679-694.

[32] Heckman, James (1979). "Sample Selection Bias as a Specification Error." *Econometrica*. 47(1): 153-161.

[33] Hendel, Igal and Alessandro Lizzeri (1999). "Adverse Selection in Durable Goods Markets." *American Economic Review*. 89(5): 1097-1115.

[34] Hendel, Igal and Alessandro Lizzeri (2002). "The Role of Leasing under Adverse Selection." *Journal of Political Economy*. 110(1): 113-143.

[35] Hendel, Igal, Alessandro Lizzeri, and Marciano Siniscalchi (2005). "Efficient Sorting in a Dynamic Adverse-Selection Model," *Review of Economic Studies.* 72(2): 467-497.

[36] Ichimura, Hidehiko (1993). "Semiparametric Least Squares and Weighted SLS Estimation of Single Index Models." *Journal of Econometrics.* 58: 71–120.

[37] Ichimura, Hidehiko and Lung-Fei Lee (1991). "Semiparametric Least Squares Estimation of Multiple Index Models: Single Equation Estimation." in (eds., Barnett, William, James Powell, and George Tauchen), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge University Press, Cambridge, 3–50.

[38] Intellichoice (2000). *The Complete Car Cost Guide 2000.* Campbell, California.

[39] Janssen, Maarten and Santanu Roy (2002). "Dynamic Trading in a Durable Good Market with Asymmetric Information." *International Economic Review*. 43(1): 257-282.

[40] Johnson, Justin and Michael Waldman (2010). "Leasing, Lemons, and Moral Hazard." *Journal of Law and Economics*. 53(2): 307-328.

[41] Khan, Shakeeb and Denis Nekipelov (2012). "Information Structure and Statistical Information in Discrete Response Models." Unpublished manuscript.

[42] Krasnokutskaya, Elena (2012). "Identification and Estimation of Auction Model with Two-Dimensional Unobserved Heterogeneity." *International Economic Review*. 53(3): 659-692.

[43] Koop, G. and D. Poirier (1997). "Learning About the Across-Regime Correlation in Switching Regression Models." *Journal of Econometrics*. 78(1): 217-227.

[44] Konishi, Hideo and Michael Sandfort (2002). "Existence of Stationary Equilibrium in the Markets for New and Used Durable Goods." *Journal of Economic Dynamics and Control*. 26(6): 1029-1052.

[45] Lacko, James. "Product Quality and Information in the Used Car Market." Staff Report. Washington: Federal Trade Commission, Bureau of Economics.

[46] Muellbauer, J. (1976). "Community Preferences and the Representative Consumer." *Econometrica*. 44(5): 979-1000.

[47] Nelson, Randy and Michael Caputo (1997). "Price Changes, Maintenance, and the Rate of Depreciation." *Review of Economics and Statistics*. 79(3): 422-430.

[48] Newey, Whitney, James Powell, and James Walker (1990). "Semiparametric Estimation of Selection Models: Some Empirical Results." *American Economic Review*. 80(2): 324–328.

[49] Olsen, Randall (1982). "Distributional Tests for Selectivity Bias and a More Robust Likelihood Estimator." *International Economic Review*. 23: 223–240.

[50] Parks, Richard (1979). "Durability, Maintenance, and the Price of Used Assets." *Economic Inquiry*. 17(2): 197-217.

[51] Peterson, Jonathan and Henry Schneider (2011). "Adverse Selection in the Used Car Market: Evidence From Purchase and Repair Patterns in the Consumer Expenditure Survey." Unpublished manuscript.

[52] Pickrell, Don and Paul Schimek (1999). "Growth in Motor Vehicle Ownership and Use: Evidence from the Nationwide Personal Transportation Survey." *Journal of Transportation and Statistics*. 2(1): 1-17.

[53] Rust, John (1985). "Stationary Equilibrium in a Market for Durable Assets." *Econometrica*. 53(4): 783-805.

[54] Rust, John (1987). "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica.* 55(5): 999-1033.

[55] Rothwell, Geoffrey and John Rust (1997). "On the Optimal Lifetime of Nuclear Power Plants Author(s): On the Optimal Lifetime of Nuclear Power Plants." *Journal of Business & Economic Statistics*. 15(2): 195-208.

[56] Schiraldi, Pasquale (2011). "Automobile Replacement: a Dynamic Structural Approach." *Rand Journal of Economics*. 42(2): 266–291.

[57] Schmalensee, Richard (1974). "Market Structure, Durability, and Maintenance Effort." *Review of Economic Studies*. 41(2): 277-287.

[58] Stern, Steven (1989). "Estimating a Simultaneous Search Model." *Journal of Labor Economics*. 7(3): 348-369.

[59] Stern, Steven (1991). "Imputing a Continuous Income Variable from a Bracketed Income Variable with Special Attention to Missing Observations." *Economics Letters*. 37(3): 287-291.

[60] Stern, Steven (1996). "Semiparametric Estimates of the Supply and Demand Effects of Disability on Labor Force Participation." *Journal of Econometrics*. 71(1): 49-70.

[61] U.S. Census Bureau (1995). *Statistical Abstract of the United States: 1995 (114th Edition).* Washington, DC.

[62] U.S. Census Bureau (2000). *Statistical Abstract of the United States: 2000 (119th Edition).* Washington, DC.

[63] Ward's Communications. *Ward's Automotive Yearbook*. Southfield, MI, various editions.

[64] Ward's Communications (2000). *Ward's Motor Vehicle Facts and Figures.* Southfield, MI.