JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **24**: 1–33 (2009)
Published online 24 October 2008 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jae.1034

ANNUAL MILES DRIVE USED CAR PRICES

MAXIM ENGERS,[a] MONICA HARTMANN[b] AND STEVEN STERN[a]*

[a] *Department of Economics, University of Virginia, Charlottesville, Virginia, USA*
[b] *Department of Economics, University of St Thomas, St Paul, Minnesota, USA*

SUMMARY

This paper investigates whether the net benefits from owning a vehicle, proxied by annual miles driven, explain the price declines observed over a vehicle's life. We first model the household decision on how much to drive each of its vehicles. Then we empirically establish that variation in household annual miles across brands explains observed price declines. Furthermore, the effect of vehicle age on annual miles decisions (and consequently on market value) depends on household characteristics and the composition of the vehicle stock owned. Copyright © 2008 John Wiley & Sons, Ltd.

Received 29 August 2005; Revised 25 December 2007

1. INTRODUCTION

The net flow of benefits provided by a vehicle can be viewed as the value of transportation services provided minus the maintenance and repair costs incurred. Suppose that the annual mileage serves as a proxy for this flow. Then annual miles may explain the observed drop in used car prices over a vehicle's working life. This paper investigates what factors influence household miles decisions and whether these choices are consistent with the observed pricing paths of these assets.

There are a number of alternative explanations for why used car prices decline with age—the pricing path we observe. One explanation is obsolescence due to styling changes or quality improvements such as airbags—advancements that are available in newer cars. For example, Purohit (1992) finds that the rate of depreciation on small vehicles increases with the introduction of a newer model with greater horsepower. However, this obsolescence effect, although significant, is small relative to other durable goods that have more radical model changes. Thus, it seems unlikely that improvements in car quality or in styling changes can explain the magnitude of price declines and the variation in price declines over brands. Another explanation is that the adverse selection problem increases as vehicles age.[1] For instance, Engers *et al.* (2008) provide evidence that the used car market for 4-year-old vehicles suffers more from adverse selection than the 1-year-old market. Still, it is questionable whether this is sufficiently large to drive the decline in prices.

To determine whether variation in vehicle usage (i.e., annual miles) helps explain the observed price declines for automobiles, we first model the household vehicle annual miles decision and

* Correspondence to: Steven Stern, Department of Economics, University of Virginia, Charlottesville, VA 22904, USA. E-mail: sns5r@virginia.edu

[1] One may argue that because of the vast improvements in car quality adverse selection is no longer a serious problem. According to J. D. Power & Associates' Survey of Initial Car Quality, the average number of car problems per 100 vehicles has declined significantly in the last 7 years. Despite these improvements, there is still variation across brands (e.g., Lexus 81, Mazda 149, Suzuki 151). Furthermore, these figures provide no indication about the variability within a brand to make a statement that the adverse selection problem has weakened.

then establish the empirical relationship between annual miles and used car prices. To estimate the household-vehicle annual miles decision, we estimate two models: a basic OLS model and a structural model. The OLS model assumes a linear relationship between household characteristics and annual miles. In contrast, the structural model allows for nonlinear influence of changes in car portfolios and household demographic characteristics on how a vehicle's age affects annual miles. For instance, it takes into account how the portfolio of vehicles owned—the number and the age distribution of the vehicles—influences a household's decision to drive a car on a particular trip and how much it is driven. Our results indicate that the structural estimates of the household mileage decision better explain the observed pattern of price decline over a vehicle's life than the OLS model. Furthermore, the structural model also allows us to decompose the age effect of annual mileage on prices into three parts: a direct aging effect, a household–car portfolio effect, and a household demographics effect. The first component captures the direct effect vehicle age has on annual miles. The portfolio effect reflects the fact that annual miles depend on the characteristics of the stock of vehicles the household owns. For example, households often choose to drive the newer, more dependable vehicle on long trips. Finally, the third effect measures the influence household characteristics, such as the number of drivers and whether or not the household lives in an urban area, have on driving patterns. The superiority of the structural model suggests that one cannot estimate the effect age has on annual miles decisions independent of household characteristics and household–car portfolio choices. Thus, studies that measure the impact of adverse selection on prices, for instance, miss these feedback effects if they use only annual miles to control for exogenously depreciated quality. They must also control for difference in driving patterns across households, vehicle age, and brand.

Our model of vehicle utilization and used car prices differs from previous studies. Goldberg (1998), using micro-level data, models annual miles as a function of household demographic and vehicle characteristics as well as characteristics of the household's stock of vehicles (e.g., number of vehicles, average age of the stock, age of the youngest vehicle). Our study of vehicle utilization differs in three ways. First, we examine driving patterns over a vehicle's entire working life and not just in its first year. Second, we investigate whether driving patterns are constant across brands. Quality deterioration is known to vary across brands. Thus, the value of the transportation services may vary by brand and lead to different patterns of car usage over the vehicle's working life. Third, we examine whether the amount a given vehicle is driven depends on its quality relative to all of the other vehicles owned by the household and not just the average quality of the household's vehicle stock.

Our model of the process that generates used car prices differs from Alberini *et al.* (1995) (AHM1) and Alberini *et al.* (1998) (AHM2) as well. They estimate the distribution of vehicle values based on what owners would accept to scrap their vehicles.[2] Because their sample includes only vehicles that qualify for a vehicle scrappage program, the average age is 17 years with a minimum and maximum age of 13.5 and 32 years, respectively. Thus, they estimate the distribution of vehicle values for only the right tail of the age distribution. In addition to expanding the range of ages analyzed, our model allows annual miles to play a different role in determining vehicle values. AHM1 and AHM2 model household and vehicle characteristics directly determining the vehicle's value. Instead, we model these characteristics affecting the vehicle's value only indirectly through

[2] The amount a household is willing to accept to scrap its car is weakly tied to the car's net benefits of transportation services. Alternatively, we use Kelly Blue Book used car prices to proxy the car's value, which captures both supply and demand market conditions.

DOI: 10.1002/jae

annual miles. Unlike AHM1 and AHM2, we find that household characteristics are determinants of a vehicle's value, but through their influence on annual miles. The intuition behind our result is that vehicle and household characteristics affects a household's driving patterns and consequently affects the vehicle's value as well.

Our finding that vehicle age affects annual miles nonlinearly also has implications for other automobile-related studies. For example, both Goldberg (1998) and Verboven (2002) assume annual miles are constant over the vehicle's life. Goldberg (1998) uses a model of vehicle utilization demand to estimate elasticity of annual miles demand with respect to gasoline cost per mile. Substitution patterns derived from this model take into account differences in fuel efficiency across new and used vehicles. Because it does not allow for differences in vehicle usage across age, the model overestimates annual miles demand for older vehicles and the reverse for newer ones, potentially biasing elasticity estimates.

Several data sources are used to establish the relationship between annual miles and used car prices. Data from the National Personal Transportation Survey allow us to track how much households drive each vehicle they own. The survey also provides information on demographic characteristics that influence the amount of driving. Detailed pricing information is collected from Kelly Blue Books. Finally, observed scrappage rates by brand are used to estimate jointly the parameters of the process that determines used car prices and scrappage rates, correcting for a potential selection bias problem in the price estimates.

The layout of the paper is as follows. Section 2 considers a simple dynamic model of the relationship between annual miles and price to motivate the remainder of the paper. Section 3 describes the annual miles data and then presents two alternative models to estimate household annual miles decisions. Section 4 presents the model that jointly estimates prices and scrappage rates to control for a selection bias. Section 5 examines the relationship between annual miles and price and demonstrates that the structural model of household annual miles decisions best explains the price decline observed in the used car market. Section 6 concludes.

2. DYNAMIC MODEL OF ANNUAL MILES AND PRICE

The empirical results of this paper show that variation across brands in rates of decline in annual miles explains variation in price declines across brands. In order to understand these results, consider the following simple model of annual miles and prices. Suppose that the benefit flow net of costs derived from owning a car of brand b and age t is denoted by f_{bt} which monotonically decreases with t. We think of annual miles, the flow of miles driven in one year, as a proxy for f_{bt}. Throughout the paper, we are interested in the flow of miles driven in one year rather than in the stock of miles previously driven. For simplicity, suppose that the rate of decline in f_{bt} is uniform over time (but can vary from brand to brand), so that

$$f_{bt} = \zeta_b^f + \xi_b^f t \tag{1}$$

where $\zeta_b^f > 0$ and $\xi_b^f < 0$.

If β is the per-period discount factor, the value W_{bt} to any consumer of a car of brand b of age t satisfies a simple recursion condition: the car can either be sold at price p_{bt} or kept for the period, generating current net surplus of f_{bt}. Thus

$$W_{bt} = \max\{f_{bt} + \beta W_{bt+1},\ p_{bt}\} \tag{2}$$

DOI: 10.1002/jae

Competition among buyers and the possibility of free disposal ensure that

$$p_{bt} = \max\{W_{bt}, 0\} \tag{3}$$

Because $f_{b0} > 0$ and f_{bt} decreases at a constant rate, there is a (unique) final period t_b^* in which the vehicle provides a positive service flow:

$$\exists t_b^* : p_{bt_b^*} = W_{bt_b^*} > 0,\ p_{bt_b^*+k} = W_{bt_b^*+k} = 0\ \forall k > 0 \tag{4}$$

Let ΔW_{bs} denote the change in W_{bs} when s is increased by one, so that $\Delta W_{bs} = W_{bs+1} - W_{bs}$. In all periods $t \le t_b^*$, $W_{bt} = f_{bt} + \beta W_{bt+1}$, and so

$$\Delta W_{bt} = \xi_b^f + \beta \Delta W_{bt+1}\ \forall t < t_b^* \tag{5}$$

By induction on $t_b^* - t$, it follows that

$$\Delta W_{bt} < 0\ \forall t < t_b^* \tag{6}$$

so prices decline monotonically until the car becomes worthless. Differentiating gives

$$\frac{\partial \Delta W_{bt+1}}{\partial \xi_b^f} = 1 + \beta \frac{\partial \Delta W_{bt+1}}{\partial \xi_b^f}\ \forall t < t_b^* \tag{7}$$

and, again by induction on $t_b^* - t$,

$$\frac{\partial \Delta W_{bt+1}}{\partial \xi_b^f} > 0\ \forall t < t_b^* \tag{8}$$

This last relationship says that if we compare two different brands then the one whose annual miles decline at a faster rate (i.e., whose ξ_b^f is a larger negative number) will also be the one whose price declines faster. The goal of this paper is to see whether this relationship between the annual miles changes and price changes is found in the data we examine.

3. ANNUAL MILES

This section presents factors that influence driving patterns. We first examine relationships in the raw data between annual miles and household and vehicle characteristics. Then a simple OLS regression of log annual miles on these characteristics provides insight into the influence of each factor on driving patterns. Finally, we present a structural model that links annual miles choices with household characteristics and the quality of the vehicles the household owns.

3.1. Data

The annual miles data for this study come from the Nationwide Personal Transportation Survey, 1995 (NPTS). The NPTS is a sample of 47,293 vehicles owned by 24,814 households. We observe

DOI: 10.1002/jae

some characteristics of each household in the sample such as family income, location characteristics (e.g., urban), and the number of drivers disaggregated by age, gender, and work status. Also, we observe each vehicle owned by the household, its age, brand, and annual miles driven.

Table I reports moments of the data. The average household has a log income of 5.93 ($37,600)[3] with 1.85 drivers and 1.91 vehicles. It has 0.2 teenage drivers and 0.45 drivers who do not work, and it drives 23,462 miles per year. The average vehicle in the sample is 9.91 years old.

Figure 1 presents broad brand categories which will be disaggregated further for most of the analysis. The categories included in the analysis are GM (Buick, Chevrolet, Geo, Oldsmobile, Pontiac, Saturn); Japanese (Honda, Mazda, Mitsubishi, Nissan, Subaru, Toyota); Ford (Ford, Mercury); Luxury (American Luxury,[4] European Luxury,[5] Japanese Luxury[6]); Chrysler (Chrysler, Dodge, Plymouth); European (Volkswagen, Volvo); Truck;[7] and Other.[8] Trucks are the most common category. Our method of aggregation is somewhat different from others in the literature. Goldberg (1998) aggregates into categories of ‘subcompact’, ‘compact’, ‘intermediate’, ‘standard’, ‘luxury’, ‘sports’, ‘pick-up truck’, and ‘van’. We feel our aggregation scheme is better for our purposes because we think there are more likely to be brand effects in the annual miles decision. In fact, Goldberg’s results suggest that variability in car quality based on car size is not a statistically significant predictor of annual miles decisions.[9] Thus, given that we are disaggregating by brand, not disaggregating further by model causes a problem only to the extent that the mileage decision is affected by variation in car quality within a brand that is not correlated with size. Berry *et al*. (1995) use a much finer aggregation scheme. Using the level of aggregation in their work would lead to imprecise estimates in our work, especially given some of the sample sizes of complementary data used in related work (Engers *et al*., 2004, 2008).

Table I. Sample moments

Variable	Mean	SD	Variable	Mean	SD
Household characteristics			# Working drivers	1.40	0.86
Log (Income)	5.93	0.70	# Vehicles	1.91	0.82
Urban	0.61	0.49	Annual miles (100)	234.62	172.89
# Drivers	1.85	0.72			
# Adult drivers	1.65	0.62	*Vehicle Characteristics*		
# Male drivers	0.92	0.54	Age	9.91	5.97

Notes:
1. The number of households is 24,814, and the number of vehicles is 47,293.
2. Income is measured in $100 units.
3. Annual miles are measured in 100-mile units.

[3] Income is measured in $100 units.
[4] Lincoln and Cadillac are aggregated throughout the analysis.
[5] Audi, BMW, Jaguar, Mercedes-Benz, Porsche, and Saab are aggregated thoughout the analysis.
[6] Infiniti and Lexus are aggregated throughout the analysis.
[7] ‘Truck’ includes all vehicles that are not automobiles.
[8] The most common brands in ‘Other’ are ‘not reported’, Hyundai, and Eagle.
[9] Furthermore, there is no standard definition for vehicle size. Choo and Mokhtarian (2004) report alternative vehicle classification schemes used in the academic literature and government agency reports. These studies primarily group vehicles based either on size, function (e.g., sedan, coupe, truck), or both. In addition to no standard definition, the classifications are not constant across time and thus somewhat arbitrary. For example, Consumer Reports’ ‘midsize’ category has changed over time, particularly during the periods when the ‘compact’ category was eliminated (1980–1983 and 1995–present).

DOI: 10.1002/jae

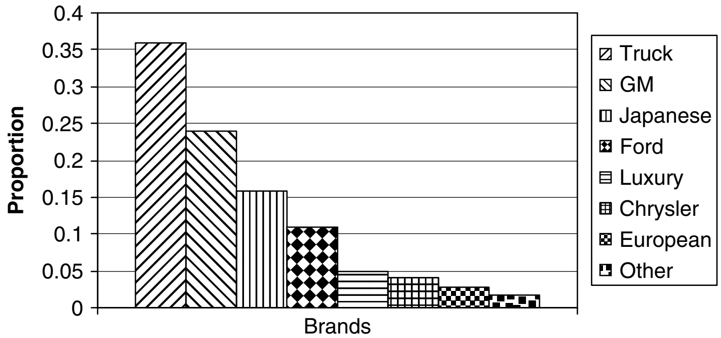

Figure 1. Vehicle composition by brand

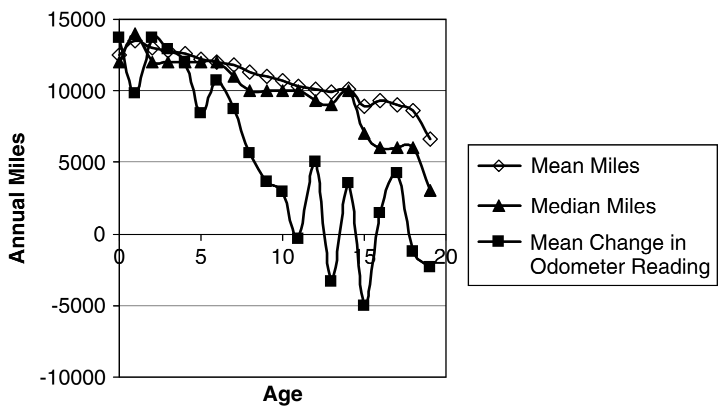

Figure 2. Annual miles driven by vehicle age

There are three measures of annual miles in the sample. We observe answers to questions of the form: 'About how many miles was this vehicle driven?' The time unit is a year if the vehicle was purchased more than a year ago, and it is the tenure of the car otherwise. In either case, an annualized miles variable is created. Also, we observe the odometer reading at the time of the interview and a few weeks after the interview. Pickrell and Schimek (1999) report results using reported annual miles and imputed annual miles based on the change in the odometer (after adjusting for season, etc.) Their estimates look quite similar to each other. Using the odometer reading, we also can construct a measure of annual miles as a function of age by aggregating vehicles into cells by age and brand and then differencing the annual miles readings. Figure 2 shows how different estimates of annual miles vary with the age of the vehicle. The estimate based on odometer reading is very imprecise and, in fact, for older vehicles, is frequently negative. For the remainder of the analysis, we focus on reported annual miles.[10]

[10] We find that the odometer reliability problem is not confined to the NPTS. Engers *et al.* (2004) look at odometer readings and observe similar problems in the CEX. There, 8% of the observations for which one is able to calculate quarterly miles using odometer readings give negative results.

DOI: 10.1002/jae

Figure 2 shows a marked decline in annual miles as a function of age.[11] This is consistent with previous work (Pickrell and Schimek, 1999). There are many outliers in the annual miles data, so we checked to see if there are large differences between mean and median annual miles. While there are some differences, especially at late ages, these are not substantial enough to warrant using orthogonal regression methods. One reason for the decline in annual miles with age is that the vehicle becomes less useful and the household drives the vehicle less. Lave (1994) suggests an alternative in which household heterogeneity causes some households to drive new cars many miles and then sell them and other households to drive fewer miles and hold the cars longer. Rust (1985) and Hendel and Lizzeri (1999) also have models of household heterogeneity with similar results. Later, in Section 5, we show that the household effects discussed in these papers are important in understanding the relationship between changes in annual miles and changes in used car prices. In any case, older vehicles are driven less than newer vehicles.

We can disaggregate the data by brand and then plot annual miles as a function of age by brand. One can see significant variation in annual miles (negative) growth rates by brand. Figure 3 graphs the intercepts and slopes of log-linear regression lines for each brand. One sees significant variation in both base (annual miles of new cars) and annual miles growth rates. The three brands with positive growth rates are Saturn (0.066), Japanese Luxury (0.049), and Geo (0.033). All three of these cases have truncated age distributions and are thus not that informative. However, even if these three are treated as outliers, the remaining range in growth rates is still quite substantial, varying from −0.094 (for Other) to −0.021 (for Volvo). It is also clear that growth rate and base are negatively correlated.

Household characteristics also might affect annual miles. The most obvious household characteristic is the number of drivers. Figure 4 shows the distribution of annual miles, disaggregated by the number of drivers in the household. Among households with a single driver, 58% drive less than 15,000 miles per year, and 98% drive less than 40,000 miles per year. Among households with two drivers, only 28% drive less than 15,000 miles, and 88% drive less than 40,000 miles. There is a smaller increase in annual miles for households with three drivers and then essentially

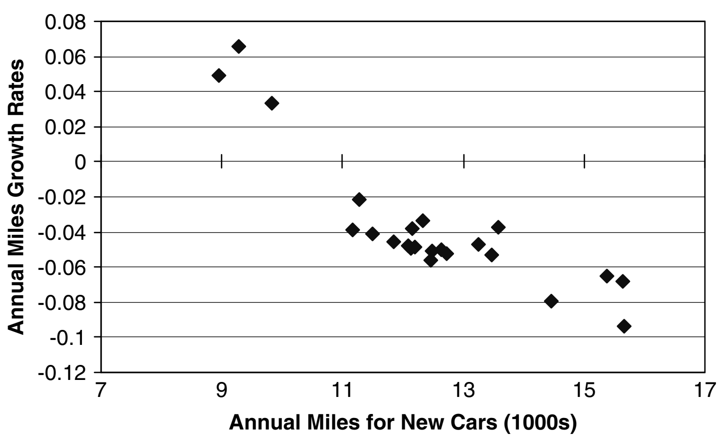

Figure 3. Intercepts and slopes of log-linear regressions between annual miles and vehicle age for each brand

[11] We cannot control for variation in model year at the same time as age of the vehicle because we have only a cross-section.

DOI: 10.1002/jae

no increase for households with more drivers. This may suggest that (a) there are economies of scale in driving[12] or that (b) not all drivers are equal. Later we characterize drivers by their age, gender, and work status.

Figure 5 is consistent with both (a) and (b) above. It shows that, while vehicle ownership increases as household size increases from one to two, further increases in household size have very little effect on the number of vehicles in the household. In all cases with household size greater than one, the modal number of vehicles is two.

We can also observe how miles across vehicles within a household portfolio vary with the number of vehicles in the household. Figure 6 provides such information. In households with two vehicles, the best car[13] is driven 65% of the time (height of the first bar for rank 1), and the second best is driven 35% of the time (height of the first bar for rank 2). In households with three

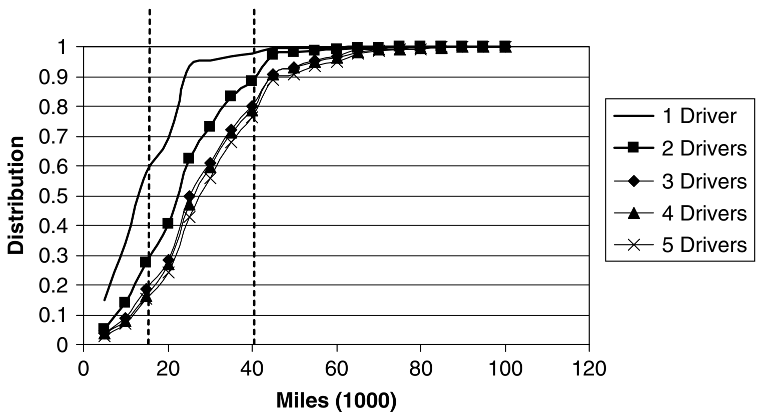

Figure 4. Distribution of annual miles driven by number of household drivers

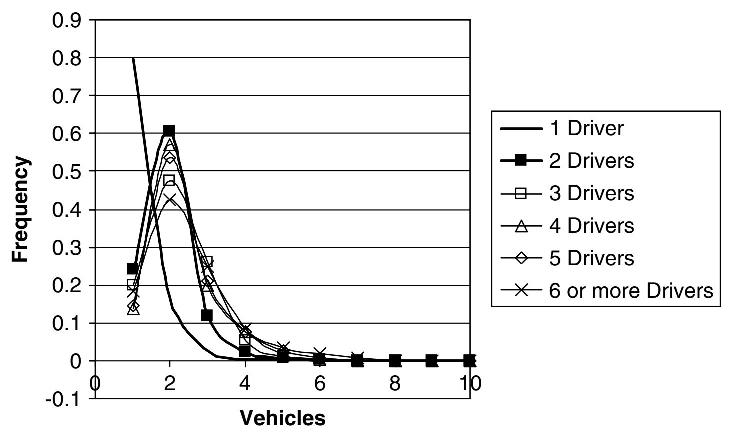

Figure 5. Frequency of vehicles by number of household drivers

[12] For example, 58% of single-driver households drive less than 15,000 miles, but 83% of two-driver households drive less than 30,000 miles, and 91% of three-driver households drive less than 45,000 miles.
[13] The best car is the one that is driven the most.

DOI: 10.1002/jae

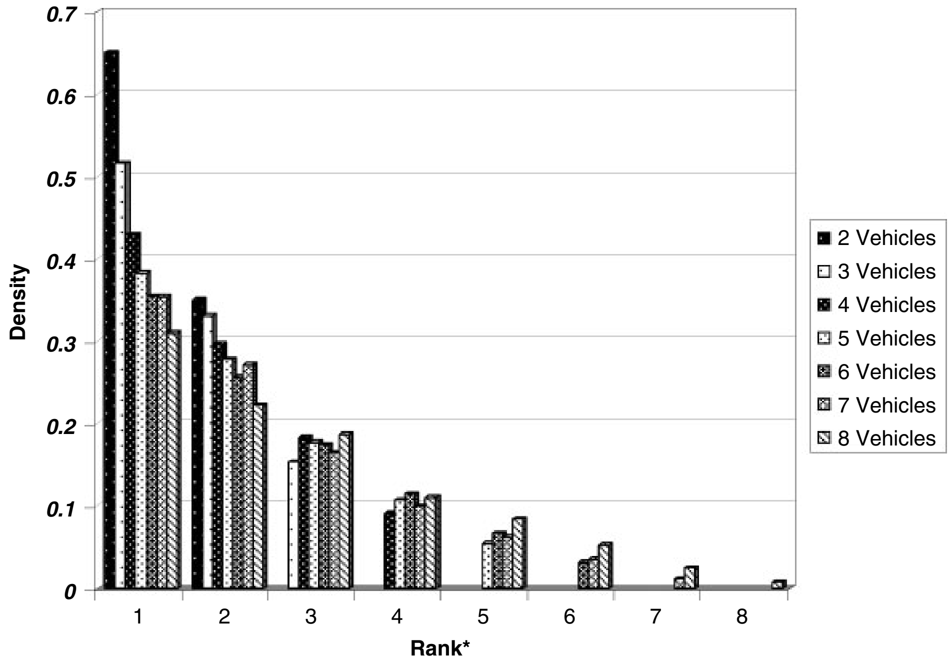

Figure 6. Density of annual miles across vehicles by number of vehicles. * Vehicle rank within household portfolio (1 = best, ...)

vehicles, the best is driven 52% of the time (height of the second bar for rank 1), the second best is driven 33% of the time (height of the second bar for rank 2), and the third best is driven 15% of the time. It is clear that, in households with more than three vehicles, those beyond the third best have little effect on the usage of the first three best (each cluster flattens out after three vehicles). This result is used later in the structural estimation of miles.

3.2. Nonstructural Estimates

In order to begin to understand what explains driving patterns, we first run an OLS regression of log annual miles on vehicle and household characteristics, ignoring the nonlinear relationships suggested by Figure 4. The specification of the estimated model is

$$\log m_{ij} = \lambda_{00} + \lambda_{10} a_{ij} + \sum_{k=1}^{K} \lambda_{0k} b_{ijk} + \sum_{k=1}^{K} \lambda_{1k} b_{ijk} a_{ij} + \sum_{k=1}^{K} \lambda_{2k} b_{ijk} \min(a_{ij}, 5) + \sum_{h=1}^{H} \lambda_{3h} D_{ih} + e_i + \varepsilon_{ij}, \quad e_i \sim \text{i.i.d.}(0, \sigma_h^2),\ \varepsilon_{ij} \sim \text{i.i.d.}(0, \sigma_v^2) \tag{9}$$

DOI: 10.1002/jae

where m_{ij} is annual miles of vehicle j in household i, b_{ijk} is a dummy for whether vehicle j is brand k, a_{ij} is the age of vehicle j, and D_i is a vector of H household characteristics.[14] Though OLS is used to estimate $(\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, standard errors are corrected to account for household effects captured by e_i.[15] The results are reported in Table II.

Table II. Log annual mileage OLS regression estimates

	Brand dummy (λ_{0k})	Brand × Age(λ_{1k})	Brand × min[Age, 5](λ_{2k})
	Vehicle characteristics		
Constant $(\lambda_{00}, \lambda_{10})$	4.595**	−0.099**	
	(0.078)	(0.006)	
Buick	−0.202**	0.054**	−0.011
	(0.108)	(0.009)	(0.023)
Chevrolet	−0.303**	0.031**	0.049**
	(0.084)	(0.007)	(0.016)
Chrysler	−0.312**	0.003	0.080**
	(0.132)	(0.013)	(0.038)
Dodge	−0.268**	0.006	0.067**
	(0.101)	(0.009)	(0.024)
Ford	−0.284**	0.032**	0.045**
	(0.079)	(0.007)	(0.014)
Geo	−0.055	0.052	−0.016
	(0.212)	(0.082)	(0.111)
Honda	−0.381**	0.055**	0.057**
	(0.085)	(0.010)	(0.021)
Mazda	−0.438**	0.055**	0.054
	(0.133)	(0.017)	(0.041)
Mercury	−0.389**	0.028**	0.070**
	(0.109)	(0.010)	(0.027)
Mitsubishi	−0.222	0.058	0.028
	(0.154)	(0.044)	(0.073)
Nissan	−0.299**	0.029**	0.056**
	(0.100)	(0.011)	(0.027)
Oldsmobile	−0.305**	0.041**	0.042*
	(0.110)	(0.008)	(0.025)
Plymouth	−0.437**	0.008	0.106**
	(0.151)	(0.011)	(0.036)
Pontiac	−0.368**	0.028**	0.058**
	(0.095)	(0.008)	(0.022)
Saturn	−0.400**	−0.715	0.877
	(0.122)	(0.988)	(0.993)
Subaru	−0.099	0.019	0.041
	(0.148)	(0.018)	(0.047)
Toyota	−0.380**	0.045	0.053**
	(0.091)	(0.010)	(0.022)
Volkswagen	−0.275*	0.033**	0.065*
	(0.163)	(0.009)	(0.038)
Volvo	−0.620**	0.051**	0.104**
	(0.180)	(0.014)	(0.050)

(*Continued*)

[14] The inclusion of the min $(a_{ij}, 5)$ term allows for a spline in age with a node at 5 years.
[15] The variance parameters, σ_h^2 and σ_v^2, are estimated by matching second sample moments to theoretical moments implied by σ_h^2 and σ_v^2.

DOI: 10.1002/jae

Table II. (*Continued*)

	Brand dummy (λ_{0k})	Brand × Age(λ_{1k})	Brand × min[Age, 5](λ_{2k})
	Vehicle characteristics		
American Luxury	−0.316**	0.031**	0.064**
	(0.120)	(0.009)	(0.029)
European Luxury	−0.539**	0.020**	0.130**
	(0.193)	(0.009)	(0.034)
Japanese Luxury	−0.538	0.084	0.059
	(0.192)	(0.342)	(0.365)
Truck	−0.273**	0.027**	0.056**
	(0.065)	(0.006)	(0.007)
	Household characteristics (λ_{3k})		
# Cars	−0.195**	Urban	−0.138**
	(0.007)		(0.010)
log (Income)	0.097**		
	(0.009)		
	# Drivers	min [# Drivers, 2]	
	Household–driver characteristics (λ_{3k})		
Total drivers	−0.016	0.065**	
	(0.018)	(0.026)	
Adult drivers	0.028	−0.095**	
	(0.022)	(0.028)	
Male drivers	0.134**	−0.007	
	(0.038)	(0.040)	
Working drivers	0.066**	0.102**	
	(0.020)	(0.022)	
Random effect standard deviations			
Household Effect (σ_h)		0.302	
Vehicle Effect (σ_v)		0.932	

Notes:
1. The number of observed households is 24,814, and the number of observed vehicles is 47,293.
2. The excluded brand is Other.
3. Numbers in parentheses are asymptotic standard errors. Asterisks indicate that items are significant at the * 10% level and at the ** 5% level.

The brand and brand/age interactions imply declining annual miles for all brands over all ages with a few exceptions: Toyota, Mazda, Volvo, and European Luxury have modest annual miles increases for the first 5 years and then significant declines in subsequent years. Saturn has large increases and then large decreases, but Saturn estimates are very imprecise. In all cases, declines in log annual miles are steeper after age 5 than before age 5.

Wald tests for H_0: $\lambda_0 = 0$ and H_0: $\lambda_1 = \lambda_2 = 0$ against the general alternatives are statistically significant at the 1% level of significance. The results also suggest that household income increases annual miles, and living in an urban area decreases annual miles. The latter result would follow from the closer proximity of places of interest and the higher cost of driving in traffic. While there are offsetting effects such as the greater abundance of places to go and higher costs of living close to work, the former effects dominate empirically. The addition of other vehicles in the household decreases annual miles on each vehicle, as Figure 6 suggests. Annual miles generally increase with drivers, especially working and adult drivers. Males drive more than

DOI: 10.1002/jae

females. A Wald test for H_0: $\lambda_3 = 0$ against H_A: $\lambda_3 \neq 0$ is statistically significant at the 1% level of significance.

3.3. Structural Estimation Method

The OLS regression reveals basic factors that influence driving patterns across households and their vehicles. This section presents a structural model that formally incorporates the nonlinear relationship between annual miles and number of drivers as evident in the data. We model annual miles choices as being determined by a two-stage process that first determines the number of trips and then determines the choice of car for each trip. The demand for trips is a function of the household's demographic characteristics and the quality of vehicles it owns. The choice of car to use for each trip is a random utility choice where the value of each mode for the particular trip has a deterministic component plus an i.i.d. Extreme Value error. This structure is the same structure as in models of brand choice such as Berry *et al.* (1995). There, one models the choice of each decision maker and then aggregates over decision makers to find shares. Here, one models the vehicle choice for each trip and then aggregates over trips to find annual miles shares for each household vehicle.[16]

Let m_{ij} be annual miles of car j in household i, $j = 1, .., J_i$, and

$$m_{i\cdot} = \sum_{j=1}^{J_i} m_{ij}. \tag{10}$$

Let household i have demographic characteristics D_i and car j of household i have characteristics C_{ij}. We can model

$$\begin{aligned} m_{ij} &= g(j, C_i, u_i, \theta) m_{i\cdot}, \\ m_{i\cdot} &= h(C_i, D_i, u_i, \theta) \end{aligned} \tag{11}$$

where $C_i = (C_{i1}, C_{i2}, .., C_{iJ_i})$, $u_i = (u_{i1}, u_{i2}, .., u_{iJ_i})$ is a vector of errors, $u_i \sim$ i.i.d.$N(0, \sigma_u^2 I)$, and θ is a vector of parameters specified below. The data suggest that, even for households with four or more cars, the characteristics of only the three best cars matter in the h function (see Figure 6).[17]

Let the value of car j be[18]

$$V_{ij} = C_{ij}\alpha + u_{ij} \tag{12}$$

[16] Technically, we must also assume that all trips are for one mile. However, since we have no data on trip distances, it is both unlikely we could identify the parameters of a richer specification and unlikely that the violation of our assumption affects results in a significant way.

[17] Figure 4 shows that the marginal effect of drivers after the second driver is small also. However, we use characteristics of all of the drivers because it is easy to do so and it is not obvious how to rank drivers.

[18] One should note that we do not allow the value of car j to depend upon the household's demographics. Berry *et al.* (1995) found valuation of vehicle characteristics (e.g., miles per gallon, size) varied across income groups, and one might expect these vehicle characteristics to influence how much a vehicle of a particular brand is driven. For instance, Hondas with high mpg may be driven over longer distances than Volvos with low mpg. While these are good arguments, since we use only age and brand as car characteristics, we think it unlikely we could measure household demographic/car characteristic interactions with any precision.

DOI: 10.1002/jae

and let $R_i(j)$ be the rank of car j in household i according to V_{ij}; i.e., $R_i(j) < R_i(k)$ iff $V_{ij} > V_{ik}$ and $R_i(j) \in \{1, 2, .., J_i\}$.[19] Let $S(\cdot)$ be the inverse of $R_i(\cdot)$.

We model h semiparametrically using polynomials with imposed monotonicity. In particular, define $p_0(v) \equiv 1$ and

$$p_j(v) = \sum_{m=1}^{M} \delta_{jm} v^m \tag{13}$$

as an Mth-degree polynomial in v for $j > 0$. Let

$$\Delta V_{iS(j)} = \begin{cases} 0 & \text{if} j = 0 \\ V_{iS(j)} & \text{if} j = 1 \\ V_{iS(j)} - V_{iS(j-1)} & \text{if} j > 1 \end{cases} \tag{14}$$

i.e., $\Delta V_{iS(1)}$ is the value of i's best choice, and $\Delta V_{iS(j)}$ is the negative of the marginal reduction in value between the jth best choice and the $(j-1)$th best choice for $j > 1$. Define $J_i^* = \min(J_i, 3)$ and

$$\Phi[J_i^*, V_i] = \sum_{j=0}^{J_i^*} \sum_{k=j}^{J_i^*} \kappa_{jk} p_j(\Delta V_{iS(j)}) p_k(\Delta V_{iS(k)}) \tag{15}$$

as a product of polynomials in different combinations of $\Delta V_{iS(j)}$ and $\Delta V_{iS(k)}$ with $\kappa_{00} = 0$ and $\kappa_{01} = 1$. We think of $\Phi[J_i^*, V_i]$ as a semiparametric approximation of a general function in $(\Delta V_{iS(j)})_{j=0}^{3}$ in the sense that, as the sample size increases, we can increase M (and also the number of ranked cars in the function from 3). This point is formalized in the Appendix.

Note that $\Phi[J_i^*, V_i]$ is equivalent to a function in $(V_{iS(j)})_{j=0}^{J_i^*}$ that can be constructed by expanding equation (15) and then collecting terms appropriately. The specification in equation (15) is preferred because it makes it easier to impose conditions such as $\partial\Phi[J_i^*, V_i]/\partial\Delta V_{iS(j)}$ declining in j.[20] Consider the special case where $\kappa_{jk} = 0$ for all $j > 0$ and only linear terms $(= 1)$ occur in equation (13); i.e.,

$$\Phi[J_i^*, V_i] = \sum_{k=j}^{J_i^*} \kappa_{0k} \Delta V_{iS(k)} \tag{16}$$

Then, for example, if all car values increase by 1, $\Phi[J_i^*, V_i]$ increases by 1; and, if just the best car increases in value by one, then $\Phi[J_i^*, V_i]$ increases by $1 - \kappa_{02}$. All of the remaining δ_{jm} terms in equation (13) and κ_{jk} terms in equation (15) allow for nonlinearity. The δ_{jm} terms allow $\Delta V_{iS(k)}$ to enter $\Phi[J_i^*, V_i]$ nonlinearly, and the κ_{jk} terms allow for interactions among the $p_j(\Delta V_{iS(j)})$ terms.[21]

We could model $h(C_i, D_i, u_i, \theta)$ as a function of $\Phi[J_i^*, V_i]$ and D_i, but such a specification would allow for the possibility that $\Phi[J_i^*, V_i]$ would have some estimated negative partial derivatives. Stern (1996) found that, without imposing monotonicity on a semiparametric function of two arguments that theory implied should be monotone, the estimated function deviated far from a

[19] We ignore ties because they occur with zero probability.

[20] We specify equation (19) in terms of $\Delta V_{iS(j)}$ rather than $V_{iS(j)}$ also because the optimization method performs significantly better. This specification implies that $\Delta V_{iS(j)} \leq 0$ for all $j > 1$.

[21] Some terms are not identified separately; e.g. the linear term in δ and the linear term in κ.

monotone function. Mukarjee and Stern (1994) and Stern (1996) suggest a simple method to impose monotonicity, similar in spirit to Matzkin (1991), and we use an analogous method here. For any function $f:W \rightarrow R$ where W is a nonempty subset of R^J for some J, define

$$\Psi[f](x) = \sup_{y \leq x}[f(y)] \tag{17}$$

Ψ transforms the function f into a new function $\Psi[f]$ that is nondecreasing in all of its arguments. In fact,

$$\Psi[f] \leq g \tag{18}$$

for all nondecreasing g such that $f \leq g$. We use the $\Psi[\cdot]$ transformation because annual miles by household should be increasing in the vector of V values. We model

$$\log h(C_i, D_i, u_i, \theta) = \Psi[\Phi] + D_i\gamma \tag{19}$$

as a partial linear equation, flexible in car values and linear in household characteristics. We approximate the Ψ function by discretizing the V-space and then checking for monotonicity in h point by point. This sounds computationally intensive, but, in fact, its cost is proportional to the number of grid points.

We assume that the g function depends on C_i and u_i only through the effect of C_i and u_i on $V_i = (V_{i1}, V_{i2}, .., V_iJ_i)$, and it depends on j only through V_{ij} and $R_{ij} = R_i(j)$. Let

$$g(j, C_i, u_i, \theta) = \frac{\exp\{\psi(V_{ij}, R_{ij}, \theta)\}}{\sum_{j'} \exp\{\psi(V_{ij'}, R_{ij'}, \theta)\}} \tag{20}$$

where

$$\psi(V_{ij}, R_{ij}, \theta) = \begin{cases} \sum_{k=R_{ij}}^{J_i-1} \omega_k(V_{iS(k)} - V_{iS(k+1)}) & \text{if } R_{ij} \leq J_i - 1 \\ 0 & \text{if } R_{ij} = J_i \end{cases} \tag{21}$$

is a spline function (in slopes) in V_{ij} with nodes at the values of the ranked V_is. This is a weighted logit function that allows for flexibility with respect to how ranking interacts with car value. Note that it is increasing in each element of V_i (assuming ω_k is positive $\forall k$) and that, if $\omega_k = 1 \forall k$, then $g(j, C_i, u_i, \theta)$ is a logit function. While there is no explicit 'outside option' in equation (20), we allow for an outside option by modeling the total demand for annual miles in equation (19). Finally, the functional forms of equations (19) through (21) satisfy the conditions described in Berry *et al.* (1995) for a unique solution in u_i to the set of equations

$$\frac{m_{ij}}{m_{i\cdot}} = g(j, C_i, u_i, \theta),\ j = 1, 2, .., J_i \tag{22}$$

$$\log m_{i\cdot} = \log h(C_i, D_i, u_i, \theta) \tag{23}$$

Define $\widehat{u}_i$ as the value of u_i that satisfies equations (22) and (23). If we decompose $\widehat{u}_i$ into $\widehat{u}_{i1}$ and $\widehat{u}_{i/1} = (\widehat{u}_{i2}, \widehat{u}_{i3}, .., \widehat{u}_iJ_i)$, then, conditional on any value of $u_{i1} = u_{i1}^0$, there is a unique solution $u_{i/1}^0$ to equation (22). Furthermore,

$$\frac{\partial u_{ij}^0}{\partial u_{i1}^0} = 1 \text{ for all} j \tag{24}$$

because of the usual location identification problem in random utility models. The value of $\widehat{u}_{i1}$ is the value that solves equation (23).

We ignore endogeneity problems. In particular, households that require a lot of annual miles (large u_i) will purchase vehicles of greater quality (larger V_{ij}). More importantly, households that like to drive (large u_i) may also like to have higher-quality vehicles (larger V_{ij}) independent of their need for annual miles.[22] Thus, the estimates of δ in equation (13) may be biased upwards. Given the nature of the data, all we can do here is acknowledge the problem and interpret results with caution.[23]

The set of parameters to estimate is $\theta = (\alpha, \gamma, \delta, \kappa, \omega, \sigma_u^2)$, where α is the effect of car characteristics C_{ij} on car values V_{ij} in equation (12), γ is the effect of demographic characteristics D_i on total household annual miles in equation (19), δ is the set of polynomial coefficients in equation (13), κ is the set of weights on each of the polynomial products used to determine the total value of the household's car portfolio in equation (15), ω is the set of weights determining how car shares depend upon relative values in equations (20) and (21), and σ_u^2 is the variance of the unobserved household heterogeneity in equation (12). The log-likelihood contribution of family i is

$$L_i = \sum_{j=1}^{J_i} \log \frac{1}{\sigma_u} \phi \left[\frac{\widehat{u}_{ij}(C_i, \theta)}{\sigma_u} \right] - \log |\Gamma(\theta)| \tag{25}$$

where $\phi[\cdot]$ is the standard normal density function and $\Gamma(\theta)$ is the Jacobian.

Our model specification differs significantly from other models of demand for annual miles. For example, Goldberg (1998) uses a model similar to Dubin and McFadden (1984) to measure the joint decision of buying a new car and how many miles to drive it. The technology available in Dubin and McFadden (1984) controls for the selection bias in annual miles associated with brand choice. As stated above, we ignore this selection bias. On the other hand, while we have a rich specification of how annual miles decisions for each car in a household's portfolio interact, Goldberg includes only aggregate characteristics of the household's portfolio (average stock age, age of newest car, whether any other cars are owned, and number of cars) and estimates a miles equation for only the new car recently purchased.[24] Mannering and Winston (1985) construct a dynamic model of car choice and utilization but also do not really allow for any interaction across cars within a household in simultaneous decisions about how much to use each car.[25]

[22] Verboven (2002) models the choice of car make, focusing on whether to buy a car with a gas or diesel engine, given desired annual miles of the household. While we ignore the endogeneity of brand choice, he ignores the endogeneity of annual miles.
[23] We cannot use the methodology developed in Verboven (2002) because our problem is more complicated. In particular, Verboven (2002) assumes each household owns only one car. While this assumption reflects European household vehicle stock in the early 1990s, the sample period for this study, a quarter of Western European households now own two or more vehicles (Ingham, 2002). This assumption is further violated in the USA, where the average household owns 1.91 vehicles. Because we are using US data and Verboven's methodology relies heavily on the one vehicle per household assumption, we had to develop another methodology to examine household annual miles decisions.
[24] Our model suggests that Goldberg's aggregated portfolio characteristics could never be a sufficient statistic for all of the characteristics of the portfolio. For example, while the number of cars matters, it must be interacted with some measure of the quality of each of the cars.
[25] In their model, the use of a car today affects the use of similar cars in the future because it affects perceptions of the car. However, there is no decision today about how to allocate driving miles among cars in the household portfolio.

DOI: 10.1002/jae

3.4. Results for Structural Specification

We estimate a number of different specifications of the structural model of the household's decision of how much to drive each vehicle it owns. The specification we focus on, reported in Table III, restricts δ in equation (13) to linear functions[26] and the higher-order shape parameters (κ in equation (19))[27] and reports estimates for ω in equation (21), car characteristic effects (α in equation (12)), demographic effects (γ in equation (19)), and log σ_u in equation (12).[28] This allows for the value of each vehicle to influence log household annual miles flexibly. We allow (in α) for brand effects and brand–age interaction effects as we did in the nonstructural model. Most of the estimates are statistically significant, and almost all of the brand–age effects are negative after adding the age slope effects.

Many of the estimates in Table III are difficult to interpret. Starting with the ω shape parameters in equation (21), recall that the number of annual miles driven on a given vehicle depends on its value relative to the other vehicles owned. Thus, ω_k captures how the relative value of the kth car to the next best vehicle affects the multinomial logit share of household miles driven on that vehicle. Consider a family with three vehicles. The log ratio of the model's choice probability for the choice with the best V to a multinomial logit model's choice probability for the choice with the best V is displayed in Figure 7 for different values of ΔV_{i2} and ΔV_{i3}. The log ratios are always positive because $\widehat{\omega}_2 > \omega_1 = 1$; this implies that values for third choices (or lower) receive less weight than values for the second choice. In particular, we can write the probability of the first choice as

$$P_1 = \frac{\exp\{\omega_1 V_1 + (\omega_2 - \omega_1) V_2\}}{\exp\{\omega_1 V_1 + (\omega_2 - \omega_1) V_2\} + \exp\{\omega_2 V_2\} + \exp\{\omega_2 V_3\}} \tag{26}$$

which converges to a multinomial logit probability as $\omega_2 \to \omega_1$ and is greater than the multinomial logit probability when $\omega_2 > \omega_1$. Also note that the choice probabilities given our specification still have the independence of irrelevant alternatives property (after conditioning on u), so Figure 7 also displays the log ratio probabilities for the second best choice.

The κ coefficients in equation (15) capture how the (total) household annual miles is dependent on the nonlinear relationships between the relative values of different combinations of vehicles from which the household takes on trips. Since we estimated only linear terms, the $h(\cdot)$ is linear as well.

The results we are most interested in are the age effects. When interpreting the age effects on the annual miles decision, consider that, for the most valuable car in a household portfolio,[29]

$$\frac{\partial \log m_{ij}}{\partial \mathrm{Age}_{ij}} = \left\{ [1 - g(j, C_i, u_i, \theta)] + \frac{1}{h(C_i, D_i, u_i, \theta)} \right\} (\alpha_{\mathrm{Age}} + \alpha_{\mathrm{Brand-Age}}), \tag{27}$$

[26] $\delta_{j1} = 1$ and $\delta_{j2} = 0$ for $j = 1, 2, 3$.

[27] $\kappa_{00} = 1$ and $\kappa_{01} = 1$ (for identification), $\kappa_{jk} = 0\ \forall_j \neq 0; \delta_1 = 1, \delta_k = 0\ \forall k \neq 1$.

[28] We estimate a third specification adding the nonlinear β terms and find that it behaves very poorly. It is clear that adding such terms leads to severe overfitting of the data, as is common in using polynomials to approximate functional form.

[29] Note that, for multinomial logit probabilities,

$$\frac{\partial \log P}{\partial A} = \frac{P(1-P)}{P}\frac{\partial P}{\partial A} = (1-P)\frac{\partial P}{\partial A}$$

DOI: 10.1002/jae

Table III. Log annual mileage structural regression estimates

	Brand dummy	Brand × Age		Brand dummy	Brand × Age
Buick	−0.038** (0.007)	0.120** (0.008)	Plymouth	0.064** (0.001)	−0.158** (0.011)
Chevrolet	0.251** (0.001)	−0.096** (0.002)	Pontiac	−0.059** (0.008)	−0.066** (0.000)
Chrysler	−0.101** (0.006)	0.164** (0.017)	Saturn	−0.032** (0.003)	−0.206 (0.027)
Dodge	−0.307** (0.002)	0.392** (0.011)	Subaru	−0.067* (0.035)	0.022 (0.041)
Ford	−0.124** (0.001)	0.361** (0.000)	Toyota	−0.032** (0.002)	0.256** (0.002)
Geo	0.584** (0.020)	0.273** (0.031)	Volkswagen	−0.076** (0.017)	0.215** (0.001)
Honda	0.161** (0.001)	0.054** (0.002)	Volvo	0.106** (0.033)	−0.045 (0.031)
Mazda	−0.202** (0.011)	−0.189** (0.010)	American Luxury	0.387** (0.003)	−0.052** (0.000)
Mercury	0.051** (0.013)	−0.076** (0.003)	European Luxury	−0.329** (0.016)	0.447** (0.013)
Mitsubishi	−0.433** (0.028)	−0.027 (0.051)	Japanese Luxury	0.131** (0.012)	−0.028 (0.116)
Nissan	0.186** (0.001)	0.063** (0.002)	Other	0.119** (0.005)	0.032 (0.007)
Oldsmobile	−0.073** (0.001)	0.130** (0.008)	Truck	−0.078** (0.000)	0.227** (0.002)
Other variables					
Constant		3.985** (0.040)	# Drivers		−0.070** (0.012)
Age		−0.337** (0.000)	# Adult drivers		0.134** (0.011)
Min (Age, 5)		0.035** (0.000)	# Male drivers		0.311** (0.012)
Log (Household income)		0.220** (0.007)	# Working drivers		0.034** (0.008)
Urban		0.093** (0.011)	$\log \sigma_u$		−0.969** (0.002)
Log-likelihood		−64967.4			
Linear semiparametric shape term					
κ_{00}		0.000	ω_1		1.000
κ_{01}		3.000	ω_2		1.456** (0.001)
κ_{02}		1.931** (0.018)	ω_3		1.149** (0.002)
κ_{03}		1.165** (0.032)	ω_4		1.246** (0.016)

Notes:
1. The excluded brand is Isuzu.
2. Numbers in parentheses are standard errors.
3. Asterisks indicate that items are statistically significant at the * 10% level and at the ** 5% level.
4. κ_{00}, κ_{01}, and ω_1 are restricted.

while, for the second most valuable car,

$$\frac{\partial \log m_{ij}}{\partial \mathrm{Age}_{ij}} = \left\{[1 - g(j, C_i, u_i, \theta)](\omega_2 - 1) + \frac{\kappa_{02}}{h(C_i, D_i, u_i, \theta)}\right\}(\alpha_{\mathrm{Age}} + \alpha_{\mathrm{Brand-Age}}), \tag{28}$$

(note that $\widehat{\omega}_2 - 1 = 0.456$ and $\widehat{\kappa}_{02} = 1.931$). However, since α_{Age} does not vary over brands ($\widehat{\alpha}_{\mathrm{Age}} = -0.302$ if Age ≤ 5 and $\widehat{\alpha}_{\mathrm{Age}} = -0.337$ if Age > 5), one can compare variation in age effects over brands by focusing just on the Brand–Age estimates. One should note that the thought experiment corresponding to $\frac{\partial \log m_{ij}}{\partial \mathrm{Age}_{ij}}$ is not to let the car age one year; rather it is to sell the car and purchase a car of the same brand one year older. The importance of this point becomes clear in Section 5.

Probably, a more fruitful way to understand age derivatives in the structural specification is to directly evaluate and graph average derivatives of log annual miles with respect to age.

DOI: 10.1002/jae

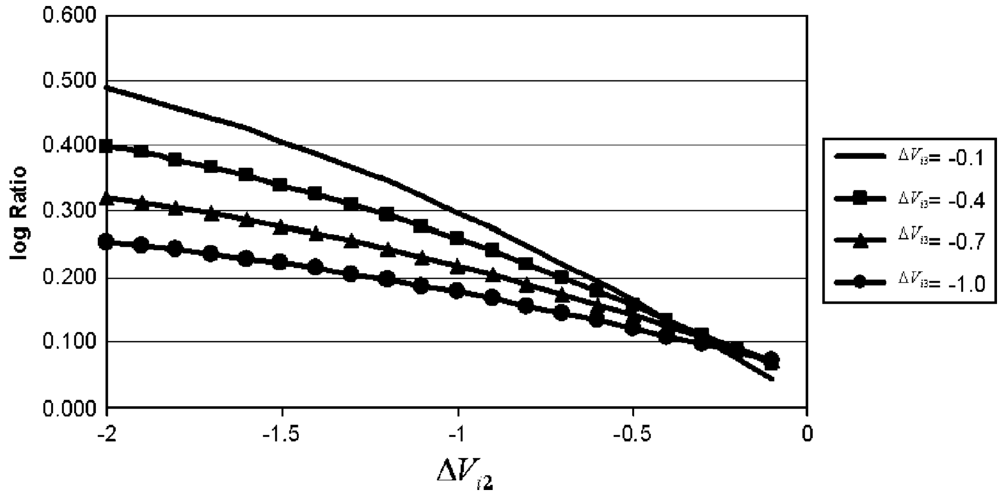

Figure 7. Log ratio of model odds to multinomial logit odds. V_{ij} is the decline in net flow benefits from switching from the $(j-1)$th best vehicle to the jth best vehicle

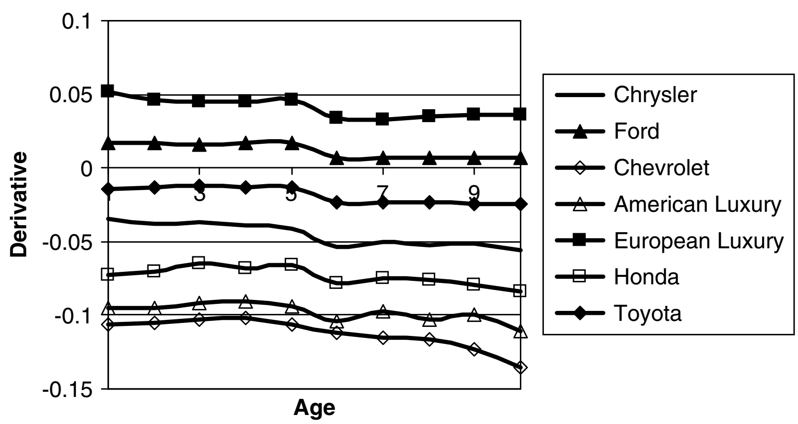

Figure 8. Average derivative of log annual miles with respect to age

Such derivatives vary by brand, age, household–car portfolios, and household characteristics. In Figure 8, we aggregate over households, therefore averaging over household–car portfolio and household characteristics effects. Figure 8 shows how average derivatives for a selected subset of brands vary with brand and age. The average derivatives are mostly negative and declining with age. Of those not represented in Figure 8, only Dodge has positive age derivatives; the other 16 are all negative and declining with age.

We can also measure how the usage of other cars in the family changes as a particular car ages. The implication of the vehicle annual miles share function in equation (20) is that, as a particular car ages (holding all else constant), it becomes less valuable (because the relevant estimated total α effects are generally negative), thus increasing the shares of the other cars in the household's portfolio. On the other hand, the household total miles function in equation (19) implies that

DOI: 10.1002/jae

total household miles decline. The structural estimates imply that the net effect is very small and negative, averaging about −0.4% per year.

We also can measure how demographic characteristics affect car usage. The estimates in Table III measure γ in equation (19), the effect of demographic characteristics on total miles. We do not interact demographic characteristics with car characteristics. Thus, the demographic derivatives do not vary over cars or over age. The results also suggest that vehicle utilization varies across households with varying numbers of drivers. Annual miles rise with the number of adult drivers, male drivers, and working drivers in the household, but fall with other types of drivers. Finally, household income increases annual miles, but unlike with the OLS estimates, living in an urban area raises annual miles.

4. PRICES

In the previous section, we show that annual miles fall as a vehicle ages. Before we can ascertain the relationship between annual miles and prices, we explore the extent to which scrappage rates explain price declines. We jointly estimate used car prices and scrappage rates as a vehicle ages. Then we examine if we need to correct for a selection bias; poorer-quality vehicles are scrapped first, biasing the estimates of price declines.

4.1. Data

The price data for this study come from the Kelley Blue Books over the period 1986–2000.[30] The price data are generated from sales reports to the National Automobile Dealer's Association (NADA) provided by member dealers. For each year in the sample period, we observe the average price of each brand of car for each relevant manufacturing year.[31] For example, in 1990 we observe the average price of multiple models for each brand for the 15 model years from 1986 to 1990 inclusive. We observe 8887 prices and 7918 changes in price. The mean price is \$10,797, and the standard deviation is \$8463. The price range is from \$825 to \$97,875. The average change in log prices is −0.133, and its standard deviation is 0.077. One can see in Figure 9 that there is a large percentage price decline in the first year and then smaller declines in subsequent years. When vehicles reach 12 years of age, price declines significantly diminish. We argue below that this can be explained by a scrapping effect. Table IV shows how price changes vary by brand.[32] There is significant variation by brand, with Honda and Toyota having the smallest average price declines and Isuzu and Hyundai having the largest average price declines. American cars tend to have larger than average price declines.

[30] See Porter and Sattler (1999) for a more detailed discussion of the quality of the Blue Book data. Alternatively, one could use transaction prices to capture car-specific quality. However, these transaction prices also reflect the relative bargaining power of the buyer and seller and other characteristics idiosyncratic to the particular car. If we were to uses transactions prices, we would average them just as Kelley Blue Book did to average out any car-specific effects.

[31] Prices based on the base four-door sedan model are used whenever possible. When matching the price data with the vehicle data in the NPTS, we assume all cars are sold at the base model price. Thus, we lose some variation in prices within a brand-model, but do capture the variation in prices across brand-models.

[32] The truck category was dropped because we lack price data. NADA records price information for trucks in a separate publication from the one that reports passenger vehicle prices used for this study. The other category was also dropped and the Isuzu and Hyundai brands were added in place because we had detailed price information on cars in this category.

DOI: 10.1002/jae

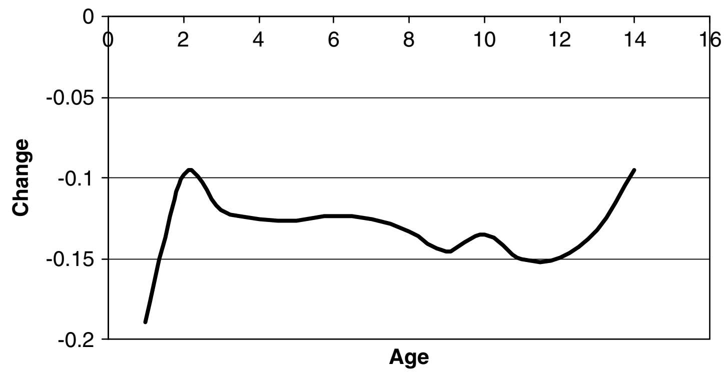

Figure 9. Change in log price by vehicle age

Table IV. Average log price decline over a vehicle's life by brand

Variable	# Obs.	Moments	Variable	# Obs.	Moments
Buick	474	−0.145** (0.065)	Nissan	395	−0.118** (0.068)
Chevrolet	675	−0.135 (0.079)	Oldsmobile	297	−0.145** (0.066)
Chrysler	282	−0.146** (0.072)	Plymouth	274	−0.135 (0.085)
Dodge	360	−0.139 (0.076)	Pontiac	503	−0.133 (0.066)
Ford	692	−0.136 (0.071)	Saturn	80	−0.121 (0.073)
Geo	65	−0.182** (0.130)	Subaru	197	−0.130 (0.099)
Honda	441	−0.097** (0.068)	Toyota	543	−0.105** (0.064)
Hyundai	117	−0.182** (0.121)	Volkswagen	121	−0.106** (0.061)
Isuzu	96	−0.176** (0.133)	Volvo	191	−0.114** (0.079)
Mazda	354	−0.120** (0.077)	American Luxury	427	−0.165** (0.067)
Mercury	338	−0.148** (0.066)	European Luxury	659	−0.131 (0.070)
Mitsubishi	181	−0.141 (0.103)	Japanese Luxury	156	−0.133 (0.064)

Notes:
1. The excluded brand is Other.
2. Numbers in parentheses are standard deviations.
3. Asterisks indicate that items are statistically significant at the ** 5% level for H_0: $\Delta \log P = -0.133$ vs. H_A: $\Delta \log P \neq -0.133$ (assuming observations are i.i.d. within a brand).

Finally, the scrapping data used in this study are from R. L. Polk & Co. The scrapping rates are observed reductions in the stock of cars of each brand as a proportion of the stock the year before. We observe scrapping rates only for eight car categories and only for a limited number of relevant vintages,[33] so we use nearby substitutes for other brands and use scrapping rates for cars manufactured in 1990. Also, we ignore the first 2 years of the car because observed scrapping rates are actually negative as car dealers sell their remaining stock. The substitution pattern is reported in Table V. One might worry that scrapping rates are not constant over time and thus it is problematic to use scrapping rates for cars manufactured in just 1990 to proxy for all scrapping rates. However, we have no data on scrapping rates disaggregated over both calendar time and brands. Also data from Ward's Communications (2001), shown in Figure 10, show that, after initial

[33] The Polk scrapping rate data were given to us by Matthew Shum.

DOI: 10.1002/jae

Table V. Scrapping rates used as proxies for other brands

Brands with observed scrapping rate	Close brand substitutes	Scrapping rate at age 4	Scrapping rate at age 9
Buick	Buick	0.0067	0.0668
Chevrolet	Chevrolet, Pontiac	0.0049	0.0832
Dodge	Dodge, Chrysler, Plymouth	0.0119	0.1072
Ford	Ford	0.0095	0.1042
Import	Honda, Mitsubishi, Mazda, Nissan, Subaru, Toyota, Volvo	0.0103	0.0653
Mercury	Mercury	0.0080	0.0969
Oldsmobile	Oldsmobile	0.0067	0.0559
Volkswagen	Volkswagen	0.0140	0.0683

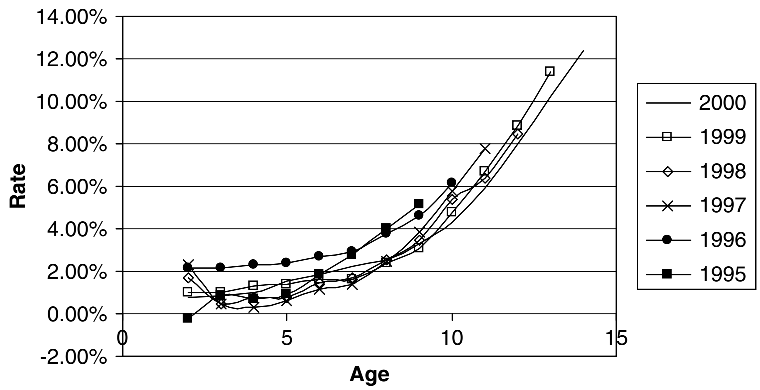

Figure 10. Scrappage rates by calendar year

volatility in scrapping rates (across calendar time), scrapping rates are relatively stable. Most of the selection effect occurs at these later ages.

4.2. Pricing Estimation Method

Given the data described above, we can estimate a pricing function for passenger vehicles controlling for scrappage. Assume the price of car j of brand i at age t is

$$p_{ijt} = \mu_{it} + \delta_{ij} \tag{29}$$

$$\delta_{ij} \sim \text{i.i.d. } N(0, \sigma^2) \tag{30}$$

with

$$\mu_{it} = \alpha_i + \kappa_i t \tag{31}$$

Let p_{it} be the average observed price of brand i cars at age t, and let d_{it} be the scrapping rate of brand i cars at age t. Scrapping occurs because some fraction d^* are totaled or otherwise exogenously disappear and because some cars have nonpositive value. We observe only positive

DOI: 10.1002/jae

prices; cars with negative prices are junked. The joint moments of (p_{it}, d_{it}) are

$$Ep_{it} = \frac{\frac{1}{\sigma\sqrt{2\pi}}\int_{-\mu_{it}}^{\infty}(\mu_{it}+\delta)\exp\left\{-\frac{\delta^2}{2\sigma^2}\right\}d\delta}{\frac{1}{\sigma\sqrt{2\pi}}\int_{-\mu_{it}}^{\infty}\exp\left\{-\frac{\delta^2}{2\sigma^2}\right\}d\delta} = \mu_{it} + \sigma\frac{\phi\left(\frac{\mu_{it}}{\sigma}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma}\right)} \quad (32)$$

$$Ed_{it} = d^* + \frac{\frac{1}{\sigma\sqrt{2\pi}}\int_{-\mu_{it-1}}^{-\mu_{it}}\exp\left\{-\frac{\delta^2}{2\sigma^2}\right\}d\delta}{\frac{1}{\sigma\sqrt{2\pi}}\int_{-\mu_{it-1}}^{\infty}\exp\left\{-\frac{\delta^2}{2\sigma^2}\right\}d\delta} = d^* + \frac{\Phi\left(\frac{\mu_{it-1}}{\sigma}\right) - \Phi\left(\frac{\mu_{it}}{\sigma}\right)}{\Phi\left(\frac{\mu_{it}}{\sigma}\right)}$$

In the data, we observe $\{p_{i1}, p_{i2}, \ldots, p_{iT}\}_{i=1}^{n}$ and $\{d_{i1}, d_{i2}, \ldots, d_{iT}\}_{i=1}^{n}$ for some subset of n car brands. Let $y_{it} = (p_{it}, d_{it})$, let $\theta = (\alpha, \kappa, \sigma, d^*)$ be the set of parameters, and consider minimizing the objective function:

$$L(\theta) = \frac{1}{nT}\sum_{i,t}(y_{it} - Ey_{it})'\Omega(y_{it} - Ey_{it}) \quad (33)$$

The asymptotic distribution of the estimates is

$$\sqrt{nT}(\widehat{\theta} - \theta) \sim N(0, \Psi) \quad (34)$$

with

$$\Psi = B^{-1}\text{plim}\left[\frac{1}{nT}\sum_{i,t}\frac{\partial E(y_{it}|\widehat{\theta})}{\partial\theta'}\cdot \Omega E((y_{it} - E(y_{it}|\theta))(y_{it} - E(y_{it}|\theta))'|\theta)'\Omega\frac{\partial E(y_{it}|\widehat{\theta})}{\partial\theta}\right]B^{-1} \quad (35)$$

where

$$B = \text{plim}\frac{1}{nT}\sum_{i,t}\frac{\partial E(y_{it}|\widehat{\theta})}{\partial\theta'}\Omega\frac{\partial E(y_{it}|\widehat{\theta})}{\partial\theta} \quad (36)$$

4.3. Pricing and Scrappage Results

The pricing and scrappage parameter estimates are reported in Table VI.[34] All of the estimates are highly significant. The most important parameters are the estimates of σ and d^*. The estimate of

[34] Additional brands were dropped because we do not have scrappage rates on their close substitutes to estimate the joint moments of (p_{it}, d_{it}). Those dropped were the three luxury brands: Geo, Hyundai, Saturn. Alternatively, one could estimate a hazard function for vehicle scrappage. Chen and Niemeier (2005) estimate retention rates by vehicle class (all passenger vehicles were in the same class). They use smog check data collected by the California Bureau of Automotive Repair. But because vehicles under age 5 are exempt from inspection, a full hazard function cannot be estimated for our purposes.

Table VI. Log price and scrapping rate declines with age by brand

	Brand effect (α_i)	Brand × age(κ_i)		Brand effect (α_i)	Brand × Age(κ_i)
Buick	13.858** (0.298)	−1.043** (0.022)	Nissan	11.222** (0.215)	−0.920** (0.017)
Chevrolet	16.230** (0.330)	−1.233** (0.024)	Oldsmobile	12.910** (0.236)	−0.995** (0.018)
Chrysler	12.616** (0.230)	−1.095** (0.020)	Plymouth	13.791** (0.249)	−1.138** (0.021)
Dodge	10.189** (0.180)	−0.916** (0.016)	Pontiac	16.239** (0.302)	−1.428** (0.026)
Ford	13.627** (0.254)	−1.115** (0.021)	Subaru	15.774** (0.347)	−1.145** (0.024)
Honda	10.740** (0.195)	−0.908** (0.016)	Toyota	14.524** (0.319)	−1.095** (0.023)
Mazda	15.152** (0.329)	−1.115** (0.023)	Volkswagon	22.350** (0.680)	−1.593** (0.047)
Mercury	16.896** (0.333)	−1.324** (0.026)	Volvo	8.231** (0.151)	−0.753** (0.014)
Mitsubishi	15.449** (0.320)	−1.175** (0.024)			
Log σ	1.345** (0.018)		$\log \frac{d^*}{1-d^*}$	−4.685** (0.007)	

Notes:
1. Prices are measured in \$1000.
2. Numbers in parentheses are standard errors.
3. Asterisks indicate that items are significant at the ** 5% level.
4. $\theta_\sigma = \log \sigma$ is estimated (rather than σ) to avoid considering negative values of σ. The estimate of σ is 3.838.
5. $\theta_d = \log \frac{d^*}{1-d^*}$ is estimated (rather than d^*) to ensure that $0 \le d^* \le 1$. Note that

$$d^* = \frac{\exp\{\theta_d\}}{1+\exp\{\theta_d\}}$$

The estimate of d^* is 0.0091.

$\log \sigma$ is 1.345, implying an estimate of σ of 3.838. While we think this may be a bit high, it is necessary to explain the curvature in the price curve and the increase in scrapping rates as cars age. The estimate of d^* is 0.0091. The estimate of d^* is very close to the observed scrapping rates of 4-year-old cars. This is what should be expected.

One can perform a residual analysis to see how well the model fits the data. Define

$$\begin{aligned} e_{it}^d &= d_{it} - E(d_{it}|\widehat{\theta}) \\ e_{it}^p &= p_{it} - E(p_{it}|\widehat{\theta}) \end{aligned} \tag{37}$$

Table VII reports average residuals for scrapping rates and prices aggregated by brand, and Table VIII reports the same aggregated by age. Though most of the average residuals are statistically significant, most are small in magnitude. Note that, for brand-specific residuals, almost all of the brands have average scrapping rate residuals and price residuals of the same sign. In cases where this is not so, then the pricing parameters (α_i, κ_i) can adjust to reduce the absolute value of both residuals. The residual analysis suggests that Nissan, Volvo, Honda, and Dodge are bad fits.[35] Part of what may be occurring, especially for Nissan, Honda, and Volvo, is that manufacturers significantly change their cars across model years during the sample period, leading to different prices paths across model years (or sets of model years). For example, the third-generation Honda Accord arrived in 1986, and Accords were upgraded when the fourth-generation Accord arrived in 1990. Thus, there may be different price paths for pre- and post-1990 Honda Accords. Because

[35] This may also be due to not controlling for how the market perceives changes in vehicle models. Purohit (1992) shows that if the market perceives a new model as less preferable than earlier models, then the demand for used versions of these vehicle rises, increasing their used prices (and vice versa). These shifts in the price paths would not be captured by our model.

DOI: 10.1002/jae

Table VII. Average residual for scrappage rates and prices aggregated by brand

Brand	# Obs.	Average residual	
		Scrappage rates	Prices
Buick	588	−0.004**	−0.032
Chevrolet	261	0.006**	−0.034
Chrysler	314	−0.006**	−0.171
Dodge	241	−0.013**	−0.203**
Ford	295	0.000	−0.115
Honda	158	−0.020**	−0.277**
Mazda	345	0.005**	0.024
Mercury	410	0.005**	−0.045
Mitsubishi	307	0.001	−0.032
Nissan	174	−0.016**	−0.236
Oldmobile	439	−0.003*	−0.129*
Plymouth	602	0.004**	0.000
Pontiac	65	−0.001	−0.012
Subaru	475	0.008**	0.057
Toyota	105	0.001	−0.006
Volkwagen	167	0.026**	0.208
Volvo	56	−0.033**	−0.474**

Note: Asterisks indicate that items are significant at the * 10% level and at the ** 5% level.

Table VIII. Average residual for scrappage rates and prices aggregated by age

Age	# Obs.	Average residual	
		Scrappage rates	Prices
3	693	0.000	0.383**
4	693	−0.004**	0.157
5	636	−0.003**	−0.082
6	574	0.011**	−0.249**
7	503	0.002**	−0.304**
8	439	0.003**	−0.242**
9	383	0.012**	−0.183**
10	325	0.017**	−0.197**
11	276	0.008**	−0.089
12	220	−0.018**	−0.142*
13	159	−0.033**	−0.084
14	101	−0.061**	−0.078

Note: Asterisks indicate that items are significant at the * 10% level and at the ** 5% level.

of this, a composition effect may be contaminating our results for these brands. It also suggests that our model is not rich enough to explain the nonlinearity in observed average pricing and its interaction with scrapping rates. In particular, we overestimate scrapping rates at ages $t \geq 12$. This is partly true because used car prices depend on many factors not controlled for in this study, such as the vehicle's equipment and condition. While these factors influence actual transaction prices,

DOI: 10.1002/jae

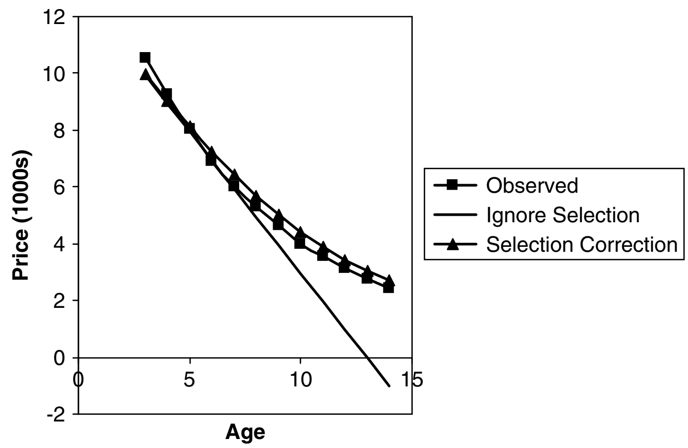

Figure 11. Price paths for Oldsmobile

they do not affect the Kelley Blue Book value. This value is just a starting point for dealers to estimate a vehicle's worth and does not take into account car-specific factors. Thus, controlling for additional factors would not improve the model's predictive power of Kelley used car prices across vehicle age.

Lastly, we examine whether one must correct our results for a selection bias. Used car prices represent the average price of automobiles still in use in time t and not the average price of vehicles in the original cohort. Failing to control for the scrapping of poorer-quality vehicles first would bias upwards the rate of decline in prices. In order to compare how alternative specifications fit the data, we can plot the observed price path, the predicted price path ignoring selection, and the predicted price path taking into account selection. For example, the price paths for Oldsmobiles are shown in Figure 11.[36] For most brands, our model is very precise in predicting the price paths after controlling for the selection bias. However, there is something about the price paths for Honda, Volvo, and Volkswagen that we miss. In particular, we cannot explain increasing price paths with our selection model when the underlying randomness is normal.

We can also plot scrapping paths.[37] We cannot explain nonmonotone scrapping rates and, in general, we miss the scrapping rates by significant amounts. This occurs basically because the terms in Ω in equation (33) associated with e_{it}^{d} and e_{it}^{p} in equation (37) give more weight to e_{it}^{p} than e_{it}^{d}.

5. RELATIONSHIP BETWEEN ANNUAL MILES AND PRICE

Now that we have estimated the process that generates both annual miles and pricing, we can investigate the relationship between these two processes. The estimates of λ from equation (9) in Table II tell us how log annual miles declines with age, disaggregated by brand, and the estimates

[36] Price paths for each brand are available at http://faculty.virginia.edu/stevenstern/resint/empiostf/selection_giffiles/paths.html.
[37] Scrapping paths for each brand are available also at http://faculty.virginia.edu/stevenstern/resint/empiostf/selection_giffiles/paths.html.

DOI: 10.1002/jae

of κ from equation (31) in Table VI tell us how long prices decline with age, disaggregated by brand.[38] We can compare annual miles and price decline estimates to see if the two processes are moving together. This is done in Figure 12. The OLS estimates of the linear relationship between them are

$$\widehat{\kappa}_j = -0.791 + 6.75\widehat{\lambda}_j + u_j \tag{38}$$

The t-statistic on the $\widehat{\lambda}_j$ coefficient is 2.00, and the R^2 is 0.21. This, along with the price/maintenance results in Engers *et al.* (2004), suggests that declines in annual miles might explain declines in price as a car ages.

Consider using the structural average derivatives with respect to age, some of which are shown in Figure 8, to explain the change in log prices. The estimated slope is 0.557, and $R^2 = 0.557$, and $R^2 = 0.020$. So the obvious question is why the OLS estimates explain prices pretty well and the structural estimates do not. The answer is that the two sets of estimates are measuring very different objects. In particular, the structural estimates are measuring $\frac{\partial \log \text{Miles}}{\partial \text{Age}}|_{\text{Demographics,Portfolios}}$; i.e., unlike the OLS estimates, they are holding constant demographic effects and household–car portfolio effects. We can interpret these differences in light of the model of car turnover in Rust (1985) or the base model in Hendel and Lizzeri (1999) (with no transactions costs or asymmetric information problems). In these models, because there are no transactions costs or asymmetric information problems, each household sells its cars every period. Changes in prices as the car ages reflect both $\frac{\partial \log \text{Miles}}{\partial \text{Age}}|_{\text{Demographics,Portfolios}}$ and the changes due to the movement of cars to different household types as the car ages.

We would like to measure the demographic effects and household–car portfolio effects but do not have enough data to estimate models of portfolio choice.[39] However, we can construct semiparametric estimates of total average derivatives (allowing demographic characteristics and

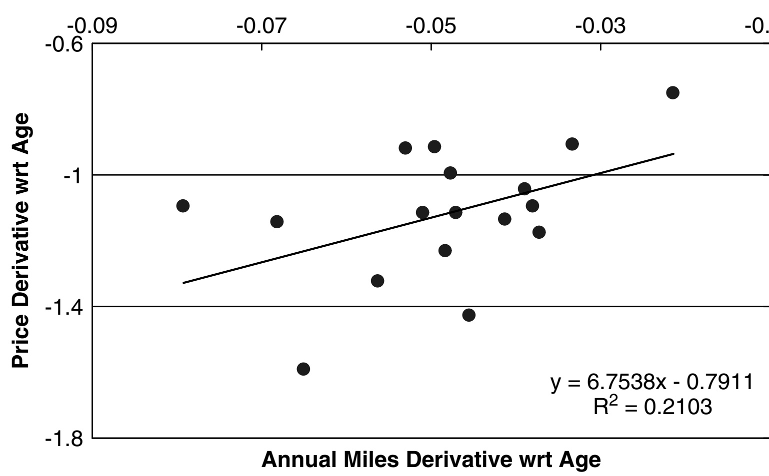

Figure 12. Relationship between OLS annual miles and price derivatives

[38] Since we aggregate car makes in the NPTS and Kelley Blue Book data the same way, we have no problem matching annual miles slope estimates from Table II with price slope estimates from Table VI.
[39] The fact that the NPTS is not a panel prevents us from measuring changes in portfolios.

DOI: 10.1002/jae

household–car portfolios to change). Using our structural estimates, we construct an estimate of the total derivative for each car j in household i in our sample. Let b_{ij} be the brand of car j in household i and a_{ij} be its age. Define $\pi_{ij} = D_i\widehat{\gamma}$ (from equation (19)) as a propensity score for household demographic characteristics in the spirit of Rubin (1979), Rosenbaum and Rubin (1983), and Heckman *et al.* (1998) and τ as a bandwidth.[40] Then our estimate of the total derivative is

$$\frac{\partial \log \widehat{\text{Miles}}_{ij}}{\partial \text{Age}} = \frac{\sum_{i'j'} 1(b_{i'j'} = b_{ij})1(a_{i'j'} = a_{ij} + 1)K\left(\frac{\pi_{i'} - \pi_i}{\tau}\right)\log m_{i'j'}}{\sum_{i'j'} 1(b_{i'j'} = b_{ij})1(a_{i'j'} = a_{ij} + 1)K\left(\frac{\pi_{i'} - \pi_i}{\tau}\right)} - \log m_{ij} \qquad (39)$$

where $1(\cdot)$ is the indicator function and $K(\cdot)$ is a kernel function. Note that we condition on cars having the same brand and being one year older. We use two separate kernel functions. First, we set $K(\cdot) \equiv 1$. This choice of kernel function causes averaging over varying demographic characteristics of households owning cars of brand b_{ij} with age equal to $a_{ij} + 1$. Then we set $K(\cdot)$ equal to a standard normal density function truncated at ± 4. This choice of kernel function controls for household demographics; i.e., we weight more heavily the annual miles of cars $i'j'$ in households with similar demographic characteristics $D_{i'}$ to household i's characteristics D_i (where similarity is measured in terms of the propensity scores).

Using the estimates from equation (39), we construct median derivatives disaggregated by brand and disaggregated by brand and age. We use median derivatives because mean derivatives are unduly influenced by outliers. In fact, Figure 13 shows how median derivatives, disaggregated by age, vary over age. The three different derivatives shown are as follows: (i) 'Age' plots the derivative, holding demographic characteristics and portfolio choices constant: (ii) 'Total' plots the derivative holding only demographic characteristics constant; while (iii) 'No Demographics' plots the derivative when both demographic characteristics and vehicle portfolio are allowed to change. Note that, while the 'Age' curve behaves well, the other two curves are poorly behaved. Curves for other brands look similar. Figure 14 presents the distribution function of derivatives (censored at ± 0.5) for Chrysler aggregated over age (other brands have similar graphs). While there is some mass in between the censoring points (in fact the median is always strictly between the censoring points), most of the mass is beyond the censoring points.

The plim of an OLS estimate in the presence of measurement error in the explanatory variable is $\lambda\sigma_x^2/(\sigma_x^2 + \sigma_e^2)$, where λ is the true value of the linear slope term, σ_x^2 is the sample variance of the explanatory variable, and σ_e^2 is the variance of the measurement error.[41] Under the assumption that a sample of random variables is i.i.d., the standard deviation of the estimate of a median is $[2\sqrt{n}f(\mu)]^{-1}$, where n is the number of observations used to estimate the median and $f(\mu)$ is the density of the random variables evaluated at the median (e.g., Bickel and Doksum, 1977, p. 400). The estimated standard deviations of the median estimates and corresponding bias ratios $(\sigma_x^2/(\sigma_x^2 + \sigma_e^2))$ are listed in Table IX. The estimated median partial derivative $\frac{\partial \log \text{Miles}}{\partial \text{Age}}|_{\text{Demographics,Portfolios}}$ is measured very precisely relative to the variation across brands σ_x. But the estimates of the total derivatives have large standard deviations.

[40] We use one quarter of the standard deviation of the propensity score as the bandwidth.

[41] In the more general case where there is heteroskedasticity in the measurement error across observations (as is true here), the term σ_e^2 is replaced by $n^{-1}\sum \sigma_{ei}^2$.

DOI: 10.1002/jae

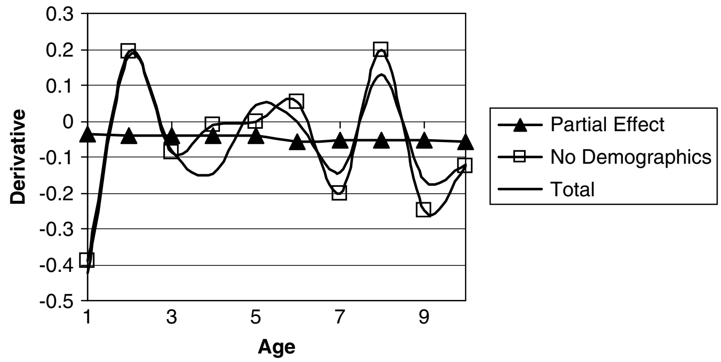

Figure 13. Effect of age on price: median log annual miles derivatives for Chrysler

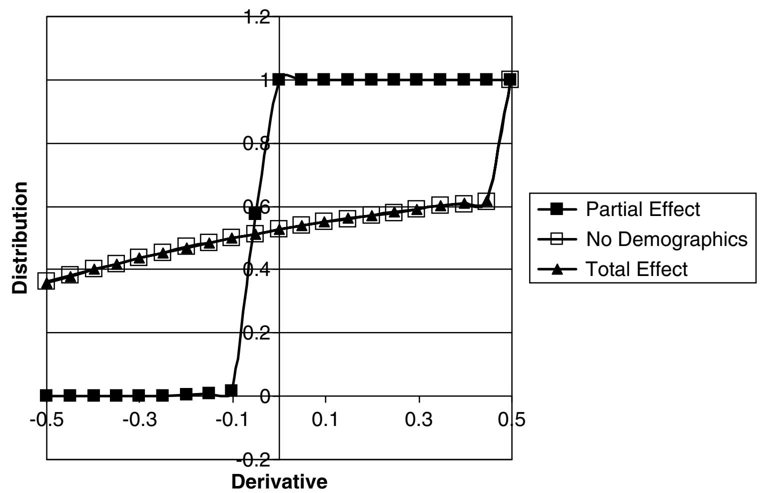

Figure 14. Effect of age on price: distribution of log annual miles derivatives for Chryler

Figures 13 and 14 and Table IX suggest that the derivatives estimated using equation (39) are problematic. However, we proceed to use them in a regression similar to equation (38). We do this partly because there is no better alternative and partly because, even though the biases of the OLS estimates are large, the R^2 statistics (and therefore t-statistics and F-statistics) have no asymptotic bias.[42] The results are presented in Table X. The first row of Table X repeats the results of

[42] This result is not as well known as the bias result. However, given the linear model

$$y = x\lambda + u$$

with

$$w = x + e,$$
$$y = wb + v,$$

DOI: 10.1002/jae

equation (38), and the second row repeats the results using the partial derivatives already discussed. The last two rows correspond to using $K(\cdot)$ equal the truncated normal density and $K(\cdot) \equiv 1$ respectively. The partial derivatives $\frac{\partial \log \text{Miles}}{\partial \text{Age}}|_{\text{Demographics,Portfolios}}$ have no explanatory power for price changes. For the linear semiparametric specification, the partial derivative is negative, casting more doubt on the credibility of the linear semiparametric estimates. On the other hand, the total derivative $\frac{\partial \log \text{Miles}}{\partial \text{Age}}$ and the derivative controlling for demographics $\frac{\partial \log \text{Miles}}{\partial \text{Age}}|_{\text{Demographics}}$ both explain price changes as well as the OLS equation from Figure 12.[43] When we further consider that the implication of Figure 14 is that median derivatives are measured with significant error (and we know that measurement error in an explanatory variable causes bias towards zero), we conclude that the structural estimates explain the price changes better than the OLS equation. These results suggest that portfolio effects and turnover effects in the spirit of Rust (1985) and Hendel and Lizzeri (1990) are important in explaining price changes as cars age.

The results also demonstrate the value of the structural estimates. While the OLS estimates in equation (38) provide a useful measure of the total effect of aging on car prices, they cannot be used to decompose the aging effect into its components: the direct aging effect, the household–car portfolio effect, and the household demographics (or car turnover) effects. The structural estimates, along with information about the distribution of cars across households, allow us to decompose the effects.

6. CONCLUSION

The objective of this paper is to determine whether changes in a vehicle's net benefits, proxied by annual miles, explain the observed decline in used car prices over the vehicle's life. We first model the household–vehicle annual miles decision using two alternative models. The OLS

Table IX. Measurement error for parametric specification

Explanatory variable	Name	σ_e	σ_x	$\frac{\sigma_x^2}{\sigma_x^2+\sigma_e^2}$
$\frac{\partial \log \text{Miles}}{\partial \text{Age}}\|_{\text{Demographics,Portfolios}}$	Partial effect of age on price	22.62	165.53	0.982
$\frac{\partial \log \text{Miles}}{\partial \text{Age}}\|_{\text{Demographics}}$	Age effect without demographics	678.95	280.45	0.146
$\frac{\partial \log \text{Miles}}{\partial \text{Age}}$	Total effect of age on price	439.80	283.15	0.293

Note: All standard deviations are multiplied by 10,000.

$$R^2 = \frac{(wb)'(wb)/n}{y'y/n} = \frac{b'(x+e)'(x+e)b/n}{(x\beta+u)'(x\beta+u)/n}$$

and

$$\text{plim } R^2 = \frac{\lambda^2\sigma_x^2}{\sigma_x^2+\sigma_e^2}\frac{\sigma_x^2+\sigma_e^2}{\lambda^2\sigma_x^2+\sigma_u^2} = \frac{\lambda^2\sigma_x^2}{\beta^2\sigma_x^2+\sigma_u^2}$$

which is independent of measurement error in x.

[43] Note that multiplying the reported estimates for the total derivative and the derivative controlling for demographics by the inverse of the associated bias terms in Table IX results in an estimate very similar to the OLS derivatives.

DOI: 10.1002/jae

Table X. Linear relationship between Δ logMiles and Δ logPrice

Explanatory variables	Estimate	
Δ logMiles (OLS)	6.754	[0.210]
$\frac{\partial \log \text{Miles}}{\partial \text{Age}}\big\|_{\text{Demographics, Portfolios}}$	0.559	[0.019]
$\frac{\partial \log \text{Miles}}{\partial \text{Age}}\big\|_{\text{Demographics}}$	2.848**	[0.196]
$\frac{\partial \log \text{Miles}}{\partial \text{Age}}$	2.957**	[0.200]

Notes:
1. Numbers in brackets are R^2 statistics.
2. Asterisks indicate that items are significant at the ** 5% level.

model restricts the relationship between annual miles and household demographics to be linear. The structural model constructed allows for a nonlinear relationship as exhibited in the data. We then compare the two models and find the structural model of household mileage decisions better explains the observed price decline in used car prices.

The intuition behind the dominance of the structural model is that it formalizes the link between annual miles choices and household and vehicle-stock characteristics. The model specification allows for car portfolios and household demographic characteristics to influence how vehicle age affects annual miles. We find that the observed decline in used car prices as a vehicle ages is best explained by decomposing the age effect into three components: the direct aging effect, the household–car portfolio effect, and the household demographics (or car turnover) effect. This implies that the effect age has on annual miles (and consequently the vehicle's value) cannot be estimated independently of household characteristics and household–car portfolio choices. The intuition behind this result is that, as a vehicle becomes older and less reliable, households are less likely to take that vehicle on long driving trips. When possible, households often take instead the newer, more reliable vehicle on those trips. The decline in usage (and subsequently annual miles) continues until the final years of a vehicle's working life where it one day primarily becomes used to transport the owner on short local trips. The estimated relationship between annual miles and vehicle value is consistent with the theoretical model of car turnover in Rust (1995) or the base model in Hendel and Lizzeri (1999) (with no transaction costs or asymmetric information problems). Portfolio and turnover effects explain price changes as a car ages.

The literature provides alternative explanations for the decline in prices: adverse selection, declining quality, and higher maintenance costs. Quality and maintenance costs both affect the net benefit flow as discussed in Section 2. While asymmetric information has been shown (empirically) to lower prices, Engers *et al*. (2004) show that this cannot be the case for maintenance costs. Thus, we are left with two competing explanations for the observed decline: adverse selection and quality deterioration.

Hendel and Lizzeri (1999) provide an empirical test to determine whether information asymmetry or exogenous depreciation (in vehicle benefits) is the dominant factor that drives price declines in the used car market. Ideally, one would like to separate out and measure each factor's effect on price declines individually. However, in order to separate these two effects, researchers must have a better understanding of how quality depreciates over the vehicle's life and how this translates

DOI: 10.1002/jae

to explaining the observed decline in used car prices as a vehicle ages. Many of these issues have been addressed in this paper. Separation of these two competing explanations for price declines is left for future work.

Our results also are relevant for policy makers who need to estimate travel demand to identify driving patterns or combat congestion, for instance. We identify the factors that determine household driving behavior. We also show that one cannot estimate annual miles just as a function of the vehicle's age and brand, but one also must factor in how household demographics and household portfolio choices influence driving decisions.

APPENDIX: DENSENESS OF POLYNOMIAL APPROXIMATIONS

Take any vector $x \in R^n$. We do a two-step change of variables, first arranging the components of x in descending order, and then taking differences. The vector $y = R(x)$, the reordering of x in descending order, consists of the same components as x but in nonincreasing order, so that $y_1 \geq y_2 \geq \ldots \geq y_i \geq \ldots \geq y_n$ and $y = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$, where σ is a permutation on the set $\{1, 2, \ldots, n\}$. If there are ties, the permutation σ is not unique, but this does not lead to a nonuniqueness in $y = R(x)$. By definition, the value of a symmetric function f is unaltered by a permutation of its arguments, and so, for all x, $f(x) = f(R(x)) = f(y)$.

Given real numbers $a < b$, any symmetric function on the box $B = [a, b]^n \subset R^n$ is completely determined by its values on the wedge

$$W = R(B) = \{y \in B : y_1 \geq y_2 \geq \ldots \geq y_i \geq \ldots \geq y_n\} \tag{A.1}$$

There is a one-to-one correspondence between the continuous symmetric functions on B and the continuous functions on W, and this one-to-one correspondence is clearly an isometry. Thus approximating a continuous symmetric function on B is equivalent to approximating a continuous function on W.

Next we introduce the transformation Δ that differences any vector in W. Each vector $y \in W$ maps to a vector $z = \Delta(y) = (y_1, y_2 - y_1, y_3 - y_2, \ldots, y_n - y_{n-1})$, and it is easy to check that the mapping Δ is one-to-one and onto the compact set

$$D = \Delta(W) = \left\{ z \in [a, b] \times [a - b, 0]^{n-1} : \sum_{j=1}^{n} z_j \in [a, b] \right\} \tag{A.2}$$

Since, for each i, $y_i = \sum_{j=1}^{i} z_j$, each polynomial in the y_i can be written as a polynomial in the associated z_i. Conversely, since each z_i is a linear combination of at most two y_i, each polynomial in the z_i can be written as a polynomial in the associated y_i. Since Δ and its inverse are linear, it is a homeomorphism between W and D.

We show that any continuous symmetric function on the box B is the uniform limit of a sequence of polynomials on W when the composite change of variable $z = \Delta \circ R(x)$ is applied to the x vector. Any symmetric function on B is completely determined by its values on W. Via the homeomorphism Δ, each continuous function on W is associated with a continuous function on D.

Proposition 1 *Any continuous symmetric function $f(x)$ on B can be uniformly approximated by a sequence of polynomials in $z = \Delta \circ R(x)$.*

DOI: 10.1002/jae

Proof We apply the Stone–Weierstrass approximation theorem (see Rudin, 1976), which states that any algebra of functions on a compact set that *vanishes nowhere* and *separates points* is dense (in the uniform norm) in the space of continuous functions on that set. The set of all polynomials on the compact set D is an algebra since the sum and product of polynomials are polynomials, as are scalar multiples of polynomials. Because (nonzero) constants are polynomials, the algebra of polynomials never vanishes at any point of D, and, given any two distinct points z', z'' in D, it is possible to find polynomials that take different values at the two points (just find an index i such that $z_i' \neq z_i''$, and then the linear polynomial z_i suffices). So the algebra of polynomials separates points. □

By the Stone–Weierstrass theorem, the polynomials on D are dense in the space of continuous functions on D with the uniform norm. Thus any continuous function on D is the uniform limit of polynomials in D.

REFERENCES

Alberini A, Harrington W, McConnell V. 1995. Determinants of participation in accelerated vehicle retirement programs. *Rand Journal of Economics* **26**: 93–112.

Alberini A, Harrington W, McConnell W. 1998. Fleet turnover and old car scrap policies. *Resources for the Future*. Discussion Paper 98–23.

Berry S, Levinsohn J, Pakes A. 1995. Automobile prices in market equilibrium. *Econometrica* **63**: 841–890.

Bickel P, Doksum K. 1977. *Mathematical Statistics: Basic Ideas and Selected Topics*. Holden-Day: San Francisco, CA.

Chen C, Niemeier D. 2005. A mass point vehicle scrappage model. *Transportation Research Part B* **39**: 401–415.

Choo S, Mokhtarian P. 2004. What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice. *Transportation Research Part A* **38**: 201–222.

Dubin J, McFadden D. 1984. An econometric analysis of residential appliance holdings and consumption. *Econometrica* **52**: 345–362.

Engers M, Hartmann M, Stern S. 2004. Automobile maintenance costs, used cars, and adverse selection. Manuscript, University of Virginia.

Engers M, Hartmann M, Stern S. 2008. Are lemons really hot potatoes? Manuscript, University of Virginia.

Goldberg P. 1998. The effects of the corporate average fuel efficiency standards in the US. *Journal of Industrial Economics* **46**: 1–33.

Heckman J, Ichimura H, Todd P. 1998. Matching as an econometric evaluation estimator. *Review of Economic Studies* **65**: 261–294.

Hendel I, Lizzeri L. 1999. Adverse selection in durable goods markets. *American Economic Review* **89**: 1097–1114.

Ingham R. 2002. Europe takes lead in tackling the car. *Agence France-Presse*, 20 September.

Lave C. 1994. What really matters is the growth rate of vehicle miles traveled. Presented at the 72nd Annual Meeting of the Transportation Research Board, Washington, DC.

Mannering F, Winston C. 1985. A dynamic empirical analysis of household vehicle ownership and utilization. *Rand Journal of Economics* **16**: 215–236.

Matzkin R. 1991. Semiparametric estimation of monotone and concave utility functions for polychotomous choice models. *Econometrica* **59**: 1315–1327.

Mukarjee H, Stern S. 1994. Feasible nonparametric estimation of multiargument monotone functions. *Journal of the American Statistical Association* **89**: 77–80.

Pickrell D, Schimek S. 1999. Growth in motor vehicle ownership and use: evidence from the nationwide personal transportation survey. *Journal of Transportation and Statistics* **2**: 1–17.

Porter R, Sattler P. 1999. Patterns of trade in the market for used durables: theory and evidence. NBER Working Paper 7149.

DOI: 10.1002/jae

Purohit D. 1992. Exploring the relationship between the markets for new and used durable goods: the case of automobiles. *Marketing Science* **11**: 154–167.
Rosenbaum P, Rubin D. 1983. The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**: 41–55.
Rubin D. 1979. Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association* **74**: 318–329.
Rudin W. 1976. *Principles of Mathematical Analysis*, 3rd edn. McGraw-Hill: New York.
Rust J. 1985. Stationary equilibrium in a market for durable assets. *Econometrica* **53**: 783–805.
Stern S. 1996. Semiparametric estimates of the supply and demand effects of disability on labor force participation. *Journal of Econometrics* **71**: 49–70.
Verboven F. 2002. Large quality-based price discrimination and tax incidence: evidence from gasoline and diesel cars. *Rand Journal of Economics* **33**: 275–297.
Ward's Communications. 2001. *Ward's Automotive Yearbook*. Southfield, MI.

DOI: 10.1002/jae