A Dynamic, Empirical Model of Adverse Selection in The Used Car Market

Maxim Engers
University of Virginia

Monica Hartmann
University of St. Thomas

Steven Stern
University of Virginia

December 2002

1 Introduction

2 Model

Let the period t utility for person i of owning car k of brand/age combination (type) j be

$$u_{ijk}^{t} = v_{ij}^{t} + \omega_{k}^{t} + \varepsilon_{ik}^{t} \tag{1}$$

where v_{ij}^{t} is a component observable by all (and possibly further decomposed into a benefit and cost term), ω_{k}^{t} is a "lemons" term observed if and only if i owns the car prior to the period, and ε_{ik}^{t} is a random component which we describe below. The value $k = 0$ denotes the option of not owning a car. We map cars into types by

$$j = \psi(k).$$

We assume (with a single exception) that, for each i and all cars k of any type j, the ε_{ik}^{t} terms are identical and can hence be denoted by ε_{ij}^{t}. This corresponds to the assumption that all cars with the same observable characteristics are treated identically by each buyer. The exception is that if i owned a car k' of type j in the previous period $t-1$ the value $\varepsilon_{ik'}^{t}$ is *not* assumed to be the same as ε_{ij}^{t}. Instead we assume that $\varepsilon_{ik'}^{t}$ is independent of all the ε_{ij}^{t} and that

$$\varepsilon_{i0}^{t}, \varepsilon_{ik'}^{t}, \varepsilon_{ij}^{t} \sim iidEV.$$

This assumption eases computation of $Emax$ terms and transition probabilities.[1] We assume that ω_{k}^{t} has a known distribution conditional on ω_{k}^{t-1},

[1] Alternatively we could assume ε varies with k and has a generalized EV structure with j determining the branching structure.

$F_\omega\left(\cdot \mid \omega_k^{t-1}\right)$.[2]

At the start of period t:

- Each person i observes ε_{ik}^t for each car k;
- The owner i of car k observes ω_k^t, but others do not. Nonowners know that the distribution of cars for sale of type j is $F_\omega^*(\cdot \mid j)$;
- Owners decide whether to sell;
- Nonowners, including those who just decided to sell, decide whether to buy, and, if so, which car;
- The market clears and trade occurs. For all j where $p_j^t > 0$, the quantity supplied (minus any exogenous supply reductions) equals the quantity demanded. If $p_j^t = 0$, the excess supply of cars, if positive, is scrapped.

The value function V is timed after all decisions have been made in a period. Let V_{ik}^t be the value to i of owning car k (of type $j = \psi(k)$) at time t, and let V_{i0}^t be the value of not owning a car. Let Ψ_i^t be the information set available to i at t. We assume that, at the horizon, $T+1$,

$$V_{ik}^{T+1} = \sum_{t=0}^{\infty} \beta^t \max\left[0, E\left(u_{ijk}^{T+1+t} \mid \Psi_i^T\right)\right].$$

In periods t before $T+1$,

$$V_{ik}^t = u_{ijk}^t + \beta E\left(M_{ik}^{t+1} \mid \Psi_i^t\right)$$

where, for all t, M_{ik}^t is the value of i's best choice in period t given i owns k at time $t-1$, and so

$$M_{ik}^t = \max_{m \in C^t}\left[E(V_{im}^t \mid \Psi_i^t) + \left(p_{\psi(k)}^t (1-\alpha) - p_{\psi(m)}^t\right) 1(m \neq k)\right]$$

where C^t is the set of choices available at t and α is an *ad valorem* transactions cost incurred by car sellers. The expectation is needed because ω_m^t is not known to i if $m \neq k$. C^t consists of the zero option of owning no car, one's current car k (if there is one), as well as a representative car of each type for sale. There is no point in considering more than one such car of a given type because i is indifferent among all such cars.

Suppose that i owned car k' at $t-1$ and owns car k at t. Then let

$$W_{ik}^t = E\left(V_{ik}^t \mid \Psi_i^{t-1}\right) + \left(p_{\psi(k')}^t (1-\alpha) - p_{\psi(k)}^t\right) 1(k' \neq k).$$

[2] In the empirical application we assume that

$$\omega_k^t = \rho\omega_k^{t-1} + \eta_k^t.$$

If $k = k'$

$$\begin{aligned} W_{ik}^t &= v_{ij}^t + E\left(\omega_k^t \mid \omega_k^{t-1}\right) + \beta E\left(M_{ik}^{t+1} \mid \Psi_i^{t-1}\right) \\ &= v_{ij}^t + E\left(\omega_k^t \mid \omega_k^{t-1}\right) + \beta \int E\left(M_{ik}^{t+1} \mid \Psi_i^t\right) dF_\omega\left(\omega_k^t \mid \omega_k^{t-1}\right) \\ &= \int V_{ik}^t dF_\omega\left(\omega_k^t \mid \omega_k^{t-1}\right) - \varepsilon_{ik}^t. \end{aligned}$$

If $k \neq k'$

$$\begin{aligned} W_{ik}^t &= v_{ij}^t + \int \left[\omega_k^t + \beta E\left(M_{ik}^{t+1} \mid \Psi_i^t\right)\right] dF_\omega^*\left(\omega_k^t \mid j\right) + \left(p_{\psi(k')}^t (1-\alpha) - p_{\psi(k)}^t\right) \\ &= \int V_{ik}^t dF_\omega^*\left(\omega_k^t \mid j\right) - \varepsilon_{ij}^t + \left(p_{\psi(k')}^t (1-\alpha) - p_{\psi(k)}^t\right). \end{aligned}$$

Now suppose that i owns k at t. Given the EV assumption about ε_{ij}^t, we can write

$$E\left(M_{ik}^{t+1} \mid \Psi_i^t\right) = \tau \ln \left[\sum_{m \in C^{t+1}} \exp\left\{W_{im}^{t+1}/\tau\right\}\right]$$

where τ is an EV variance parameter. The Williams-Daly-Zachary theorem (McFadden, 1981) generalizes away from the EV assumption and shows what properties $E\left(M_{ik}^{t+1} \mid \Psi_i^t\right)$ must have in general. Also, we can write

$$\begin{aligned} P_{im}^{t+1} &= \Pr\left[i \text{ chooses } m \text{ at } t+1 \mid \Psi_i^t\right] \\ &= \frac{\exp\left\{W_{im}^{t+1}/\tau\right\}}{\sum_{n \in C^{t+1}} \exp\left\{W_{in}^{t+1}/\tau\right\}}. \end{aligned}$$

In general, as long as the support of ε is $(-\infty, \infty)$, $0 < P_{ik}^{t+1} < 1$.

3 Model Implications

4 Data

4.1 Sales Data

We observe a random sample of 3543 cars in Virginia. For each car, we observe all sales from time 0 (9/13/93) to T (5/13/01) along with how long the owner at time 0 owned the car. Time T date is the date the Department of Motor Vehicles (DMV) of Virginia collected a random sample for us, and time 0 is the earliest date for which they had electronic records available on May 13, 2001. Thus, we observe for each observation $\{t_{ij}, v_{ij}\}_{j=1}^{J_i}$ where t_{ij} is the length of the jth ownership of car i (starting at $j = 1$ with the owner at time 0) and v_{ij} is a censoring dummy. Each record in this random sample also contains the

VIN. The VIN is a 17 digit alphanumeric string that provides information on the make and manufacture year of the vehicle and other characteristics of the car (e.g., engine size, brake system, manufacturing plant, etc.). We can deduce from this data whether the first obseved owner ($j = 1$) is the initial owner of the car by comparing the manufacture year to the date that the $j = 1$ owner purchased the car.

We have three types of ownership spells in our data. Consider a car that was manufactured in 1990, sold in 11/92, and sold again in 1/98. Our data would have an "initial spell" t_{i1} equal to the number of days[3] from 6/90 until 11/92. During this period we know only that there was a least one owner and that the owner in 11/92 sold it. Next there would be a "middle spell" t_{i2} from 11/92 to 1/98 corresponding to an observed owner, i.e., one whose ownership is observed by DMV. Finally, there would be an "end spell" t_{i3} from 1/98 to 5/01. The first spell would be censored in a unique way in that we know only that the length of the initial spell is no more than t_{i1}. It could have been less than t_{i1} because, during that period there could have been multiple owners. The second spell is uncensored, and the third is right censored in the usual way. Alternatively, consider a car that was manufactured and purchased in 1992, sold in 10/96, and then again in 8/00. Our data would have t_{i1}, the uncensored tenure of the initial owner (= # days between the new purchase date and its sale in 10/96), t_{i2}, the uncensored tenure of the owner between 10/96 and 8/00, and t_{i3}, the censored tenure of the owner between 8/00 and 5/01. In general, we define a variable d_{ij} to indicate the type of spell: $d_{ij} = 1$ if the jth spell is an "initial spell," $d_{ij} = 2$ if the jth spell is a "middle spell," and $d_{ij} = 3$ if the jth spell is an "end spell."

Table 1reports the number of spells of each type and the moments of each. There are 8584 observed spells. The average spell length is two to three years, depending on the type of spell. In our empirical section, we condition on the brand. Table 2 reports the number of observations by brand. We have aggregated Cadillac and Lincoln into "American Luxury," Audi, BMW, Jaguar, Mercedes-Benz, Porsche, and Saab into "European Luxury," Infiniti and Lexus into "Japanese Luxury," and Acura with Honda. Some small brands such as Peugot and Fiat are excluded. We include only cars that were manufactured after 1985 because Kelley Blue Book price data are available only back to 1986. The data is described in more detail in Engers, Hartmann, and Stern (2002a).

4.2 Price Data

The price data for this study comes from the Kelley Blue Books over the period 1986 - 2000. For each year in the sample period, we observe the average price of each brand of car for each relevant manufacturing year. For example, in 1990, we observe the average price of each brand from 1986 - 1990 inclusive. [add text on how "average price" is estimated] We observe

[3] When date of purchase for the initial owner is not observed, we assume arbitrarily that it is 6/30/yy where yy is the manufacture year.

8,887 prices and 7,918 changes in price. The mean price is $10,797, and the standard deviation is $8,463. 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. This data is described in more detail in Engers, Hartmann, and Stern (2002b).

4.3 Cost Data

The cost data for this study comes from the Consumer Expenditure Survey (CES). The CES, collected by the Bureau of Labor Statistics, tracks household spending habits on a variety of items including expenditures on car related items. It is a rotating panel in which each household is interviewed at most four times. Each quarter twenty percent of the sample is replaced with new households.

For each household, the survey records total household expenditures on maintenance, car insurance, gas and oil, licensing and registration fees, and other miscellaneous expenses. These figures are aggregated over all vehicles owned by the household. The survey also collects information on the stock of vehicles owned by the household, providing detailed information on each vehicle's make, model, and year. All information is collected on a quarterly basis.

Because expenditures are aggregated across household's stock of vehicles, several restrictions are placed on the sample in order to able to identify how maintenance expenditures vary by make and model of passenger vehicles. A quarterly observation for a household is included in our sample only if a) the household does not buy or sell a car in that quarter, b) the household does not own a truck, van, or motorcycle and c) each of the household's cars is manufactured in 1986 or later and is manufactured by one of the automakers listed in Table 2. Household interviews that tracked expenditures on vehicle related items between 1988 and 1999 were used. On average, households own 1.35 cars with two households owning as many as seven vehicles. Household quarterly expenditures on car related items vary from as little as zero to as high as $10,999 (in 1999 dollars). The average age of the first and second car owned by the household is 5 and 1.3 years, respectively. The average age of the remaining cars are close to zero because most households own less than two vehicles. More detail is provided in Engers, Hartmann, and Stern (2002c).

4.4 Mileage Data

The mileage data for this study comes from the Nationwide Personal Transportation Survey, 1995 (NPTS). The NPTS is a sample of 47,293 vehicles owned by 24,814 households. For each household in the sample, we observe some characteristics of the household such as the number of drivers disagregated by age, gender, and work status, family income, and location characteristics (e.g., urban). Also, we observe each vehicle owned by the household, its age, brand, and mileage. The data is described in more detail in Engers, Hartmann and Stern (2002b). Furthermore, we estimate mileage as function of personal and vehicle characteristics and use those estimates here in the flow benefits.

5 Econometric Methodology (sketch)

We observe average cost of maintenance of a brand j car at age a from the Consumer Expenditure Survey and we assume that flow benefits are proportional to mileage which is observable from the Nationwide Personal Transportation Survey. Thus, we assume we can observe (up to a proportionality constant) v_{ij}^t in equation (1). We also can get data on the survivor function for brand j cars until first sale,

$$\begin{aligned} \Psi(a,j) &= \Pr\left[\bigcup_{s\leq a} V_{ik}^t\left(\omega_k^t,\varepsilon_{ik}^t,j,s\right)\geq p(j,s)\right] \\ &= \iint\limits_{R^2\backslash S(j,a)} g_\varepsilon(\varepsilon)\, g_\omega(\omega\mid j,a)\, d\varepsilon d\omega \end{aligned} \tag{2}$$

where

$$S(j,a)=\{(\omega,\varepsilon): W\leq V(\omega,\varepsilon,j,s)\leq p+W\}$$

is the set of combinations of (ω,ε) where the individual chooses to sell the vehicle (because $V(\omega,\varepsilon,j,s)\leq p+W$) and $R^2\backslash S(j,a)=\{(\omega,\varepsilon)\notin S(b,a)\}$ is the set of combinations of (ω,ε) where the individual chooses not to sell, and $g_\varepsilon(\cdot)$ and $g_\omega(\cdot\mid j,a)$ are the densities of ε and ω respectively. The cost and mileage data identify v_{ij}^t, and *Kelley Blue Book* data identify $p(b,a)$. The remaining parameters to estimate are identified from our Virginia DMV survivor data.

6 Results

7 Conclusions

8 References

References

[1] Engers, Maxim, Monica Hartmann, and Steven Stern (2002a). "Are Lemons Really Hot Potatoes?" Unpublished manuscript.

[2] Engers, Maxim, Monica Hartmann, and Steven Stern (2002b). "Mileage and Used Car Prices." Unpublished manuscript.

[3] Engers, Maxim, Monica Hartmann, and Steven Stern (2002c). "Automobile Maintenance Costs, Used Cars, and Adverse Selection." Unpublished manuscript.

[4] McFadden, Daniel (1981). "Econometric Models of Probabilistic Choice." in C. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data with Econometric Applications*, pp 198-272. Cambridge: Cambridge University Press.

9 Tables and Figures

Table 1
Spell Moments

Spell Type	# Obs	Mean	Std. Dev.
Initial	3197	1608.0	1161.9
Middle	2379	726.0	670.0
End	3008	1164.1	865.2
Total	8584		

Note: Spells are measured in days.

Table 2
Number of Observations by Brand

Variable	# Obs		Variable	# Obs
Chrysler	82		European Luxury	114
Dodge	126		Honda	405
Plymouth	69		Mitsubishi	60
Ford	533		Mazda	154
Mercury	138		Nissan	185
Buick	145		Subaru	60
Chevrolet	346		Toyota	356
Oldsmobile	154		Volkswagen	77
Pontiac	183		Volvo	47
Saturn	51		Geo	29
American Luxury	133		Hyundai	62
Japanese Luxury	34		Total	3543

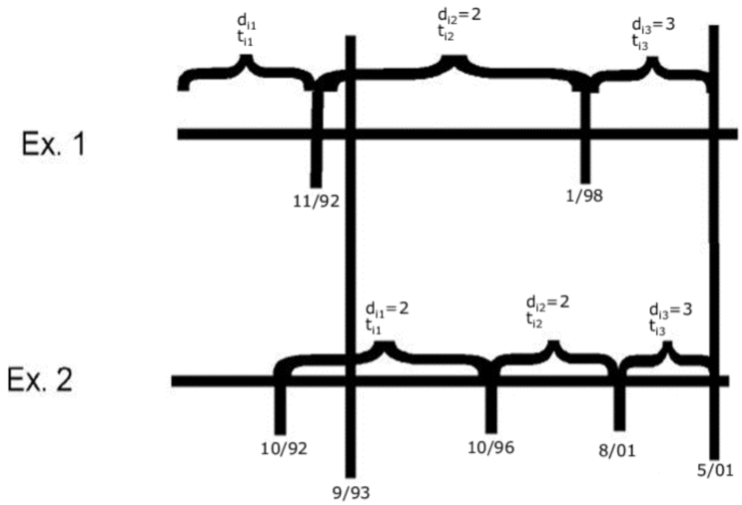

Figure 1:

Figure 1
Data Examples