Vehicle Choice Behavior and the Declining Market Share of U.S. Automakers*

by

Kenneth E. Train and Clifford Winston[1]

Forthcoming, *International Economic Review*

Abstract: We develop a consumer-level model of vehicle choice to shed light on the erosion of the U.S. automobile manufacturers' market share during the past decade. We examine the influence of vehicle attributes, brand loyalty, product line characteristics, and dealerships. We find that nearly all of the loss in market share for U.S. manufacturers can be explained by changes in basic vehicle attributes, namely: price, size, power, operating cost, transmission type, reliability, and body type. U.S. manufacturers have improved their vehicles' attributes but not as much as Japanese and European manufacturers have improved the attributes of their vehicles.

*Submitted July 15, 2005. Revision submitted February 10. 2006.

Shortened Title: Vehicle Choice

[1] We are grateful to S. Berry, F. Mannering, C. Manski, D. McFadden, A. Pakes, P. Reiss, J. Rust, M. Trajtenberg, F. Wolak, and seminar participants at Berkeley, Maryland, Stanford, UC Irvine, and Yale for helpful comments. A. Langer provided valuable research assistance.

1. Introduction

Until the energy shocks of the 1970s opened the U.S. market to foreign automakers by spurring consumer interest in small fuel-efficient cars, General Motors, Ford, and Chrysler sold nearly 9 out of every 10 new vehicles on the American road. After gaining a toehold in the U.S. market, Japanese automakers, in particular, have taken significant share from what was once justifiably called the Big Three (table 1). Today, about 40 percent of the nation's new cars and 70 percent of its light trucks are sold by U.S. producers.[2] And new competitive pressures portend additional losses in share, especially in the light truck market—a traditional stronghold for U.S. firms partly because of a 25 percent tariff on light trucks built outside of North America and the historical absence of European automakers from this market. Japanese automakers are building light trucks in the United States to avoid the tariff and introducing new minivans, SUVs, and pickups, while European automakers are starting to offer SUVs.

PUT TABLE 1 ABOUT HERE.

The domestic industry's loss in market share is not attributable to the problems experienced by any one automaker (table 2). Indeed, GM, Ford, and Chrysler are all losing market share at the same time. Toyota has recently surpassed Ford as the second largest seller of new cars in the United States and Honda has surpassed Chrysler (notwithstanding Chrysler's merger with Daimler-Benz in 1998) and is within reach of Ford. Both companies as well as Nissan (not shown) are also likely to increase their share of the light truck market as their new offerings become available. On the other hand, General Motors' share of new car and light truck sales has not been so low since the 1920s.

PUT TABLE 2 ABOUT HERE.

[2] Ford and General Motors have partial ownership of some foreign automakers. However, the industry and manufacturer shares reported here would not be affected very much if Ford's and GM's sales included, on the basis of their ownership shares, the sales of these automakers.

It may be believed that the industry's losses in share are confined to certain geographical regions of the country such as parts of the East and West Coasts and some affluent areas in the Southwest. However, Japanese and European automakers have built manufacturing plants and research and development facilities in the Midwest and Mid-south that have spurred local employment and helped increase market share in these areas because American consumers no longer view auto "imports" as costing themselves or their friends a job. In addition, during the past decade Japanese automakers in particular have significantly expanded their dealer network in interior regions of the country.

The forces that cause a tight oligopoly to lose its market dominance are central to our understanding of competition and industry performance. Academic researchers, industry analysts, and even industry executives have offered various supply-side and demand-side explanations for the U.S. automakers' decline. Aizcorbe, Winston, and Friedlaender (1987) found that Japanese automakers were able to build an additional small car during the 1970s and early 1980s for $1,300 to $2,000 less than it cost the U.S. automakers to build the same car. This cost advantage translated into greater market share for the Japanese firms. However, recent evidence compiled by Harbour and Associates suggests that the U.S.-Japanese cost differential has narrowed.[3] For example, an average GM vehicle now requires 24 hours of assembly time while an average Honda North American vehicle requires 22.3 hours. Compared with Japanese transplants, American plants have also significantly reduced the labor that they require to build a car.

Recently, industry executives such as Bill Ford of Ford and Rick Wagoner of General Motors have argued that their competitive position has been eroded by rising health care and pension costs and an undervalued yen. They have called on the federal government to provide the industry with various subsidies and tax breaks and to pressure Japan to raise the value of its currency. However, the U.S. industry's market share has been declining long before it began to incur the costs of an aging workforce and has continued to decline during times when the dollar/yen exchange rate was quite favorable for U.S. automakers.

[3] A summary is contained in *Automotive News* email alert June 2, 2005.

From a consumer's perspective, Japanese automakers have developed a reputation for building high-quality products that suggests that their technology in cars represents better value than American technology in cars. Indeed, using various measures of quality and reliability, widely-cited publications such as *Consumer Reports* and the *J.D. Power Report* have generally given their highest ratings in the past few decades to cars made by Japanese and European manufacturers rather than by American manufacturers. Changes in market share since the 1970s could therefore be explained by the relative value of the technology in domestic and foreign producers' vehicles as captured in basic vehicle attributes such as price, fuel economy, power and so on.

Consumers' preferences may also be affected by more subtle attributes of a vehicle such as the feel of a stereo knob and the shine of plastics used in interiors. Robert Lutz, General Motors' vice chairman for product development, claims that attention to these subtle attributes sends a powerful message to consumers that an automaker cares about its products.[4] An even more subtle consideration is consumers' unobserved tastes that are expressed, as John DeLorean colorfully put it, in whether their eyes light up when they walk through an automaker's showroom and whether they buy a car that they are in love with.[5] U.S. automakers may have lost market share because of the poor workmanship of their products or factors that while difficult to quantify have adversely influenced consumers' tastes toward domestic vehicles.

Brand loyalty is inextricably related to developing, maintaining, and protecting market share. Mannering and Winston (1991) found that a significant fraction of GM's loss in market share during the 1980s could be explained by the stronger brand loyalty that American consumers developed toward Japanese producers' vehicles compared with the loyalty that they had for American producers' vehicles. Ford and Chrysler were able to retain their share during that period, but the American firms' subsequent losses in share

[4] Danny Hakim, "G.M. Executive Preaches: Sweat the Smallest Details," *New York Times*, January 5, 2004.

[5] Danny Hakim, "Detroit's New Crisis Could Be its Worst," *New York Times*, March 27, 2005.

may be partly attributable to the intensity of consumer loyalty toward Japanese and European automakers.

Economic theory suggests that product line rivalry may be an important feature of competition in the passenger-vehicle market because consumers have strongly varying preferences. Industry analysts stress that it is important for automakers to develop attractive product lines that anticipate and respond quickly to changes in consumer preferences. General Motors, for example, has offered an assortment of vehicles that missed major trends such as the growth in the small-car market in the late 1970s and early 1980s, the interest in more aerodynamic midsize cars in the late 1980s, and the rise of sport utility vehicles based on pickup truck designs in the 1990s. Two key features of an automaker's product line are the range of vehicles that are offered and whether any particular vehicle generates "buzz" that spurs sales of all of the automaker's vehicles. Finally, the competitiveness of a product line is also affected by an automaker's network of dealers. Changes in market share since the 1970s could therefore reflect the relative strengths of domestic and foreign manufacturers' product lines and distribution systems.

Given the myriad of hypotheses that have been offered, it is useful to empirically assess as many of them as possible. This paper develops a model of consumer vehicle choice to investigate the major potential causes of the domestic industry's shrinking market share. A long line of research beginning with Lave and Train (1979), Manski and Sherman (1980), Mannering and Winston (1985), and Train (1986) indicates that such models are a natural way to quantify a variety of influences on consumers' behavior, some of which may prove useful for understanding the industry's decline. However, these models have accumulated several specification and estimation concerns including: the independence of irrelevant alternatives (IIA) assumption maintained by the multinomial logit model that is often used to analyze choices; the possibility that vehicle price is endogenous because it is related to unobserved vehicle attributes; the importance of accounting for heterogeneity among vehicle consumers; and the appropriate treatment of dynamic influences on choice such as brand loyalty.

We explore these concerns in the process of estimating the choices of U.S. consumers who acquired new vehicles in 2000. Although we do not claim to provide definitive solutions to all of the methodological issues that we confront, we do obtain plausible evidence that choices are strongly influenced by vehicle attributes, brand loyalty, and automobile dealerships but surprisingly they are not affected by product line characteristics. We use the choice model to simulate market shares under alternative scenarios to explore the reasons for the loss in market share by U.S. manufacturers.

We find that the U.S. industry's loss in share during the past decade can be explained almost entirely by relative changes in the most basic attributes of new vehicles, namely price, size, power, operating cost, transmission type, reliability, and body type. The result is surprising in its simplicity, implying that it is not necessary to resort to the plethora of explanations just described. Arguments based on subtle attributes such as the design of interior features, unobserved responses by consumers to vehicle offerings, or even measurable attributes beyond those listed above do not play a measurable role in the industry's competitive problems. Similarly, changes in loyalty patterns, whether an automaker's product line is broad or narrow or includes a hot car, and changes in dealership networks do not contribute much to the industry's decline. Our finding suggests that U.S. automobile executives should focus more attention on understanding why their companies seem unable to improve the basic attributes of their vehicles as rapidly as their foreign competitors are able to improve their vehicles' basic attributes, and try to remedy the situation.

2. Choice of Model and its Formulation

Our objective is to investigate the most likely determinants of market share changes in the new vehicle market during the past decade. The approach we take is to estimate the *conditional* choice of buying a new car. In a complete vehicle choice model, consumers can choose to buy a new car, buy any used car, continue using their current vehicles, or not own any vehicle and presumably rely on pubic transportation. Our model, which accounts for unobserved taste variation and is conditional on a subset of the vehicle choice alternatives (i.e., new car purchases), could yield inconsistent estimates if tastes that affect which new car the consumer chooses also affect whether the consumer chooses one of these cars instead of another alternative. It is thus useful to discuss the advantages and drawbacks of different approaches to analyzing new vehicle choices before formulating our model.

2.1 Controlling for Related Choices

One approach to the problem of related choices that is taken, for example, by Berry, Levinsohn, and Pakes (2004), is to aggregate all the other alternatives into one alternative—which is often called an outside good. The weakness of this approach is that it is difficult to specify attributes that meaningfully represent this alternative. Thus, including an outside good is still likely to yield inconsistent estimates because unobserved tastes that affect a consumer's assessment of new cars can also affect a consumer's assessment of other alternatives through the *attributes* of those alternatives. For example, the value that consumers place on vehicle price affects their evaluation of each used car based on a used car's price, not just on the existence of an unspecified outside good.[6]

A further difficulty with using an outside good is that the sample of new car buyers needs to be weighted to be consistent with the general population. These weights differ greatly over observations, because the subpopulation of new car buyers is quite different from the general population. Thus, the density of tastes among the subpopulation of new vehicle buyers is derived as being proportional to the population density times the probability of a buying a new car. But this probability is influenced by the attributes of other alternatives including but not limited to all used and currently owned vehicles. However, as noted, an outside good does not control for these attributes; hence, the conditional density is likely to be incorrectly inferred from the population density.

In our view, the distribution of preferences among new car buyers can be estimated more accurately by estimating it directly on a sample of new car buyers and by conducting extensive tests of error components that capture vehicle attributes and socioeconomic variables that are likely to affect consumers' new vehicle choices as well as their related choices. Our approach also has the practical advantage that it can include explanatory variables whose distributions are not known for the general population. In contrast, the

[6] The utility of the outside good is usually specified as a function of demographic characteristics and random terms. Although these elements tend to have significant effects, indicating that they are capturing differences between people who buy the good and those who do not, the utility of the outside good is not structural because it does not relate to the attributes of the alternatives that are subsumed into the aggregate "outside good."

outside good approach restricts the set of explanatory variables to those whose distributions in the U.S. population are known, because the population distribution is used to weight the sample. Thus, we would be precluded from exploring, among other influences, the impact of brand loyalty and an automaker's network on vehicle choice because measures of these effects are very difficult to obtain for the general population.[7]

Of course, the issues raised here could potentially be avoided by analyzing a complete model of vehicle ownership. The problems posed by this approach are cost and empirical tractability. As noted later, we must conduct a customized survey of households to collect information on such variables as past vehicle purchases, vehicles seriously considered when selecting a new vehicle, and so on. This information is not included in publicly available surveys. Customized surveys are expensive—in our case, the cost was roughly $50 per household. Households that actually acquire a new vehicle represent roughly 12 percent of the general population of households. Thus, the cost of assembling a sample of all households in the population, which would be necessary to analyze the choice of whether a consumer decides to acquire a vehicle, would run into the hundreds of thousands of dollars. For those households who actually purchase a vehicle, we would have to analyze whether they selected a new or used vehicle, which would result in an enormous choice set that could not be reduced because our model does not invoke the IIA assumption. Finally, even a complete model of vehicle ownership is open to the criticism that it is conditional on other related decisions such as mode choice to work and residential location. Using our approach as a starting point, future research can consider the trade-off between additional modeling and costly data collection and possible improvements in the accuracy of parameter estimates.

2.2 Model Formulation

Our analysis is based on a random utility function that characterizes consumers'

---

[7] By conditioning choices on the purchase of a new vehicle, we are precluded from analyzing or forecasting changes in market size. However, we are interested in decomposing potential influences on changes in market shares, especially the decline in the U.S. manufacturers' share. We can conduct this analysis without having to control for changes in market size.

choices of new vehicles by make (e.g., Toyota) and model (e.g., Camry). A mixed logit model relates this choice to the average utility of each make and model (i.e., average over consumers), the variation in utility that relates to consumers' observed characteristics, and the variation in utility that is purely random and does not relate to observed consumer characteristics. In an auxiliary regression equation, the average utility of each make and model is related to the observed attributes of the vehicle, using an estimation procedure that accounts for the possible endogeneity of vehicle prices.

We index consumers by $n = 1,...,N$, and the available makes and models of new vehicles by $j = 1,...,J$. The utility, U_{nj}, that consumer n derives from vehicle j is given by:

$$U_{nj} = \delta_j + \beta' x_{nj} + \mu_n' w_{nj} + \varepsilon_{nj}, \tag{1}$$

where δ_j is "average" utility (or, more precisely, the portion of utility that is the same for all consumers[8]), x_{nj} is a vector of consumer characteristics interacted with vehicle attributes, product line and distribution variables, and brand loyalties (capturing observed heterogeneity); β represents the mean coefficient for each of these variables in the population; w_{nj} is a vector of vehicle attributes that may be interacted with consumer characteristics (capturing unobserved heterogeneity); μ_n is a vector of random terms with zero mean that corresponds to vector elements in w_{nj}; and ε_{nj} is a random scalar that captures all remaining elements of utility provided by vehicle j to consumer n.

Brownstone and Train (1999) point out that the terms $\mu_n' w_{nj}$ represent random coefficients and/or error components. Each term in $\mu_n' w_{nj}$ is an unobserved component of utility that induces correlation and non-proportional substitution between vehicles, thus

---

[8] The explanatory variables x_{nj} have non-zero mean in general, thus average utility is actually δ_j plus the mean of $\beta' x_{nj}$. We use the term "average utility" to refer to δ_j because other terms, such as "common utility" or "fixed portion of utility," seem less intuitive. The main point is that δ_j does not vary over consumers while the other portions of utility do.

overcoming the IIA restriction imposed by the standard logit model. Note that elements of w_{nj} can correspond to an element of x_{nj}, in which case the corresponding element of β represents the average coefficient and the corresponding element of μ_n captures random variation around this average. Elements of w_{nj} that do not correspond to elements of x_{nj} can be interpreted as capturing a random coefficient with zero mean.

Denote the density of μ_n as $f(\mu \mid \sigma)$, which depends on parameters σ that represent, for example, the covariance of μ_n. Note that f is the density conditional on a new vehicle purchase and may therefore depend on observed variables in the model that arise from a consumer's optimizing behavior that leads to a new vehicle purchase. We explore the empirical form of f and its dependence on observed variables as part of our estimation.

We assume that ε_{nj} is iid extreme value. Note that the average utility associated with omitted attributes, which varies over vehicles, is absorbed into δ_j. Given the distributional assumption on ε_{nj}, the probability that consumer n chooses alternative i is given by the mixed logit model (see, e.g., Revelt and Train (1998), McFadden and Train (2000)): [9]

$$P_{ni} = \int \frac{e^{\delta_i + \beta' x_{ni} + \mu' w_{ni}}}{\sum_j e^{\delta_j + \beta' x_{nj} + \mu' w_{nj}}} f(\mu | \sigma) d\mu . \tag{2}$$

McFadden and Train (2000) demonstrate that by making an appropriate choice of variables and mixing distribution, a model taking this form can approximate any random utility model—and pattern of vehicle substitution—to any level of accuracy.

Market (or aggregate) demand is the sum of individual consumers' demand. The true (observed) share of consumers buying vehicle i is S_i. As in Berry, Levinsohn, and Pakes (2004) and Goolsbee and Petrin (2004), we use market shares rather than sample

[9] These references are for mixed logits on consumer-level choice data. Mixed logits on market share data have been estimated by Boyd and Mellman (1980), Cardell and Dunbar (1980), and more recently revived by Berry, Levinsohn, and Pakes (1995).

shares to avoid the sampling variance associated with the latter shares. The predicted share, denoted $\hat{S}_i(\theta,\delta)$, is obtained by calculating P_{ni} with parameters $\theta = \{\beta,\sigma\}$ and $\delta = \{\delta_1,\ldots,\delta_J\}$ and averaging P_{ni} over the N consumers in the sample. Berry (1994) has shown that for any value of θ, a unique δ exists such that the predicted market shares equal the actual market shares. This fact allows δ to be expressed as a function of θ, thereby reducing the number of parameters that enter the likelihood function. We denote $\delta(\theta,S)$, where $S = \{S_1,\ldots,S_J\}$, as satisfying the relation:

(3) $$S_i = \hat{S}_i(\theta,\delta(\theta,S)) = \sum_n P_{ni}(\theta,\delta(\theta,S))/N \quad i = 1,\ldots,J \,.$$

The parameters of the choice model θ are estimated by maximum likelihood procedures described below, while δ is calculated such that predicted market shares match observed market shares at θ.

The alternative-specific constant for each vehicle, $\delta_j(\theta,S)$, captures the average utility associated with observed as well as unobserved attributes, while the variables that enter the random utility model capture the variation of utility among consumers. To complete the model, we specify average utility as a function of vehicle attributes, z, with parameters, α, that do not vary over consumers:

(4) $$\delta_j(\theta,S) = \alpha' z_j + \xi_j \,,$$

where ξ_j captures the average utility associated with omitted vehicle attributes. Note that elements of w_{nj} in the random utility function given in equation (1) can correspond to an element of z_j .

Vehicle price, an element of z_j , is likely to be affected by unobserved attributes, so that ξ_j does not have a zero mean conditional on z_j . To address this problem, let y_j be a vector of instruments that includes the non-price elements of z_j plus other exogenous variables that we discuss below. The assumption that $E(\xi_j | y_j) = 0$ for all j is sufficient for the instrumental variables estimator of α to be consistent and asymptotically normal, given θ.

3. Estimation Procedures

Estimation of the random utility function presented here is complicated by our efforts to capture preference heterogeneity (i.e., σ), the average utility for each make and model (i.e., δ), and the effect of brand loyalty on vehicle choice. We discuss each of these issues in turn.

3.1 Preference Heterogeneity and Vehicles Considered

The set of vehicles that consumers consider before making a purchase provides additional information on their tastes that may be useful in identifying preference heterogeneity. We therefore asked consumers in our sample to list the vehicles that they seriously considered in addition to the vehicle that they purchased. Most consumers indicated that they considered only one vehicle besides their chosen vehicle; no consumer listed more than five vehicles.

We included this information in estimating the choice model by treating the chosen vehicle and the vehicles that were seriously considered as constituting a ranking. Consumers who indicated only one "considered" vehicle generated a utility ranking of $U_{ni} > U_{nh} > U_{nj}$ for all $j \neq i, h$ for chosen vehicle *i* and considered vehicle *h*. Consumers who indicated more than one considered vehicle generated a utility ranking in the order that they listed the vehicles.

Luce and Suppes (1965) demonstrated that when the unobserved component of utility is iid extreme value, the probability of a utility ranking, starting with the first-ranked alternative, is a product of logit formulas. Therefore, conditional on μ_n, the probability, $L_n(\mu_n)$, that a consumer buys vehicle *i* and also considered vehicle *h* is:

(5) $$L_n(\mu_n) = \left(\frac{e^{\delta_i(\theta,S)+\beta' x_{ni} + \mu_n' w_{ni}}}{\sum_{j=1}^{J} e^{\delta_j(\theta,S)+\beta' x_{nj} + \mu_n' w_{nj}}} \right) \left(\frac{e^{\delta_h(\theta,S)+\beta' x_{nh} + \mu_n' w_{nh}}}{\sum_{j=1,\, j\neq i}^{J} e^{\delta_j(\theta,S)+\beta' x_{nj} + \mu_n' w_{nj}}} \right),$$

where the sum in the second logit formula is over all vehicles except *i*. The probability of the consumer's ranking conditional on μ_n is defined analogously for consumers who listed more than one considered vehicle. The unconditional probability of the consumer's ranking is then:

(6) $$R_n = \int L_n(\mu) f(\mu|\sigma) d\mu \, .$$

We found in preliminary estimations that it was essential to include the vehicles that consumers considered to estimate the distribution of their tastes. When we included only the choice of the vehicle that consumers purchased, the parameters of the systematic part of the model were hardly affected but we were unable to obtain any statistically significant error components. In contrast, the standard deviations for several elements of μ_n were found to be significant when we included the vehicles that consumers seriously considered. Berry, Levinsohn, and Pakes (2004) also reported that they were unable to estimate unobserved taste variation without including consumers' rankings.

3.2 Average Preferences

We included dummy variables for all the makes and models in our sample to estimate consumers' average value of utility from each vehicle. In the numerical search for the maximum of the likelihood function (see below), δ is calculated for each trial value of θ. We use the contraction procedure developed by Berry, Levinsohn, and Pakes (1995) where at any given value of θ, the following formula is applied iteratively until predicted shares equal observed market shares (within a given tolerance):

(7) $$\delta_j^t(\theta, S) = \delta_j^{t-1}(\theta, S) + \ln(S_j) - \ln\left(\hat{S}_j\left(\theta, \delta^{t-1}(\theta, S)\right)\right) \quad j = 1, \ldots, J.$$

As in previous applications of this procedure, we found that the algorithm attains convergence quickly.

3.3 Brand Loyalty

Brand loyalty has been a crucial consideration in automobile demand analysis beginning with Manski and Sherman (1980), who included a transactions dummy variable in their vehicle choice model, Mannering and Winston (1985), who included lagged utilization variables, and Mannering and Winston (1991), who included "brand loyalty" variables defined as the number of previous consecutive purchases from the same manufacturer. We use the last measure of brand loyalty here. The notion of brand loyalty suggests that households may behave myopically with respect to their vehicle ownership decisions—that is, they do not take full account of the impact of their present consumption of automobiles on future tastes. Indeed, households do appear to behave myopically as indicated by high implicit discount rates based on vehicle purchase decisions (Mannering and Winston (1985)) and by frequent breaks in loyalty. Accordingly, researchers have not modeled consumers' vehicle choices as arising from the maximization of an intertemporal utility function subject to an intertemporal budget constraint.

We specify separate brand loyalty variables in our model for GM, Ford, Chrysler, Japanese manufacturers as a group, European manufacturers as a group, and Korean manufacturers as a group. However, care must be taken when interpreting these coefficients (Mannering and Winston (1991)). One interpretation, which is based on the idea of state dependence that we are attempting to capture, posits that a consumer's ownership experience with a manufacturer's products builds confidence in that manufacturer (e.g., reduces perceived risk) thereby producing a greater likelihood that a consumer will buy the manufacturer's products in the future. Consumers' actual experiences with a manufacturer's vehicles determine the intensity of their loyalty—positive experiences are reflected in a large coefficient for the manufacturer's loyalty variable. An alternative interpretation is that the loyalty variable captures unobserved taste heterogeneity among consumers that is not controlled for elsewhere in the model: previous purchases reflect consumers' tastes that influence their current purchase.

As Heckman (1991) pointed out, state dependence and consumer heterogeneity are fundamentally indistinguishable unless one imposes some structure on the way observed and unobserved variables interact. In our case, we contend that it is more likely that brand loyalty is capturing state dependence rather than heterogeneity because it is defined for manufacturers that produce a wide range of vehicles, especially when Japanese and European vehicles are each considered as a group. Unobserved heterogeneity is more likely to be associated with makes and models than with manufacturers. For example, if a middle-aged male bought a Honda S2000 in the past because it best matched his tastes, then, based on his revealed tastes, it is reasonable to expect that he would be more likely to buy a Porsche Boxer or a Mercedes SLK in his current choice than to buy a Honda Accord or Toyota Camry.

Our brand loyalty variables could nevertheless be subject to endogeneity bias to the extent that they relate to unobserved tastes for vehicle attributes; that is, the distribution of random terms in the choice model may be different conditional on different values of the brand loyalty variables. Heckman (1981a,b) pioneered the development of dynamic discrete choice models with lagged dependent variables and serially correlated errors, recognizing the critical role of initial conditions. However, applying his methods to address the possible bias of brand loyalty coefficients is thwarted by formidable data and computational requirements. First, we would have to obtain data for all sampled consumers indicating their vehicle choices and the attributes of the vehicles that were available at the time of each previous purchase *beginning with the first vehicle that they ever purchased*. Second, we would have to simultaneously estimate previous and current vehicle choice probabilities incorporating these data and a plausible specification of how consumers' tastes are likely to change over time.

We therefore take a simpler and more tractable approach that while not necessarily leading to a consistent stochastic structure, can be expected to capture the primary differences in the error distribution of the random utility function conditional on our brand loyalty variables. As reported later, we also estimate the model without any loyalty variables and find that the estimates for all other parameters are nearly the same with and without the loyalty variables. Hence, any inconsistency that is induced by the loyalty variables and our treatment of the conditional error distribution is confined to the loyalty parameters themselves and does not affect other parameters.

We represent the information contained in the loyalty variables about consumers' preferences across manufacturers by denoting each consumer's manufacturer preference as η_{nm}, with $m = 1,...,6$ indexing the six manufacturer groups (GM, Ford, Chrysler, Japanese, European, and Korean.) These preferences result from the manufacturers' offerings and consumers' tastes for the vehicles' attributes. In the past, consumer n chose the manufacturer with the highest value of η_{nm}. The unconditional distribution of $\eta_n = \{\eta_{n1}, \ldots, \eta_{n6}\}$ is $g(\eta_n)$. The distribution of η_n conditional on the consumer having chosen manufacturer m is:

(8) $$h\left(\eta_n \middle| \eta_{nm} > \eta_{ns} \forall\ s \neq m\right) = \frac{I\left(\eta_{nm} > \eta_{ns}\ \forall\ s \neq m\right) g\left(\eta_n\right)}{\int I\left(\eta_{nm} > \eta_{ns}\ \forall\ s \neq m\right) g\left(\eta_n\right) d\eta_n},$$

where $I(\cdot)$ is a 0-1 indicator of whether the statement in parentheses is true.

For the current choice, the utility of vehicle j, which is produced by manufacturer $s(j)$, is as previously specified plus a term $\lambda\eta_{ns}$, where λ is the coefficient of the additional element of utility. Conditional on the past choice of manufacturer, the choice probability is then the logit formula with this term added to its argument, integrated over the conditional density of η_n. Formally, the probability that consumer n chooses vehicle i produced by manufacturer $s(i)$, given that the consumer chose a vehicle by manufacturer m in the past (where m may equal $s(i)$) is:

(9) $$P_{ni} = \iint \frac{e^{\delta_i + \beta' x_{ni} + \mu_n' w_{ni} + \lambda\eta_{ns(i)}}}{\sum_{j=1}^{J} e^{\delta_j + \beta' x_{nj} + \mu' w_{nj} + \lambda\eta_{ns(j)}}} f\left(\mu \middle| \sigma\right) h\left(\eta_n \middle| \eta_{nm} > \eta_{ns}\ \forall s \neq m\right) d\mu d\eta_n .$$

This choice probability is a mixed logit with an extra error component whose distribution is conditioned on the consumer's past choice of manufacturer. Similarly, the probability for the observed choices of consumer n, who for instance bought vehicle i and ranked vehicle h as second, is the same as equation (5) but expanded to include the extra error component:

(10) $$R_n = \iint \frac{e^{\delta_i + \beta' x_{ni} + \mu_n' w_{ni} + \lambda\eta_{ns(i)}}}{\sum_{j=1}^{J} e^{\delta_j + \beta' x_{nj} + \mu' w_{nj} + \lambda\eta_{ns(j)}}} \frac{e^{\delta_h + \beta' x_{nh} + \mu_n' w_{nh} + \lambda\eta_{ns(h)}}}{\sum_{j\neq i, j=1}^{J} e^{\delta_j + \beta' x_{nj} + \mu' w_{nj} + \lambda\eta_{ns(j)}}} f\left(\mu \middle| \sigma\right) h\left(\eta_n \middle| \eta_{nm} > \eta_{ns}\ \forall s \neq m\right) d\mu d\eta_n \ .$$

Note we also account for additional ranked choices as appropriate.

3.4 Estimators

The choice probabilities, P_{ni}, in equation *(9)* and the ranking probabilities, R_n, in equation *(10)*, are integrals with no closed form solution. We use simulation to approximate the integrals. The simulated choice probability is:

(11) $$\tilde{P}_{ni} = \frac{1}{D}\sum_{d=1}^{D}\frac{e^{\delta_i(\theta,S)+\beta' x_{ni}+\mu_d' w_{ni}+\lambda\eta_{rns(i)}}}{\sum_j e^{\delta_j(\theta,S)+\beta' x_{nj}+\mu_d' w_{nj}+\lambda\eta_{rns(j)}}}\ ,$$

for draws μ_d, $d = 1,\ldots,D$ from density $f(\mu \mid \sigma)$ and draws from the conditional distribution *h*. The probability of consumer *n's* purchased and ranked vehicles are simulated similarly, giving $\tilde{R}_n$.

The simulated log-likelihood function for the observed first and ranked choices in the sample is $LL = \sum_n \ln \tilde{R}_n$, which is maximized with respect to parameters $\theta = \{\beta,\sigma\}$ and λ. As described above, estimates of $\delta = \{\delta_1,\ldots,\delta_J\}$ are obtained using the iteration formula in equation (7) to ensure that predicted shares equal observed market shares.[10] Goolsbee and Petrin (2004) also use maximum likelihood procedures to estimate choice probabilities. Petrin (2002) and Berry, Levinsohn, and Pakes (2004) used a generalized method of moments estimator with moments based on consumer-level choices.

We use 200 Halton draws for simulation.[11] Halton draws are a type of low-discrepancy sequence that, as *R* rises, has coverage properties that are superior to pseudo-random draws. For example, Bhat (2001) and Train (2000) found that 100 Halton

---

[10] Our sample size is small relative to the number of available makes and models, and thus relative to the number of elements in δ. However, this is not problematic because observed market shares rather than sample shares are used to determine δ. Note that the sample of new vehicle buyers is large relative to the number of elements in θ that reflect differences in preferences among households, and it is this sample that is used to estimate θ.

[11] Draws from the conditional distribution *h* were obtained by an accept/reject procedure: draw values of η_n from $g(\eta_n)$ and retain those for which $\eta_{nm} > \eta_{ns}$ for all $s \neq m$. We assume $g(\eta_n)$ is a product of standard normal variables and use 200 accepted draws in the simulation of the integral over η_n.

draws achieved greater accuracy in mixed logit estimations than 1000 pseudo-random draws.[12] To estimate the impact of different numbers of draws on parameter estimates, we estimated the model using 100, 150, and 200 draws. The estimates differed an average of 8 percent when we increased the number of draws from 100 to 150 and differed an average of 4 percent when we increased the number of draws from 150 to 200. The differences are well within the confidence intervals for the parameters and indicate that simulation noise and bias are sufficiently small to not warrant further increases in the number of draws. In addition, we evaluated the log-likelihood function, gradient and Hessian using 400 draws at the parameter estimates obtained with 200 draws. The average log-likehood changed only very slightly, from -6.5163 to -6.5141. The test statistic $g'H^{-1}g$, where g is the gradient vector and H is the Hessian, took the value 0.00351. Under the null hypothesis that the gradient is zero, this test statistic is distributed chi-squared with degrees of freedom equal to the number of parameters. The extremely low value indicates that we cannot reject the hypothesis that the gradient using 400 draws is zero at the estimates using 200 draws at any meaningful level of significance. For these reasons, we concluded that using 200 Halton draws for simulation was sufficient. We report robust standard errors that take into account simulation noise, as suggested by McFadden and Train (2000).

After estimating the ranked choice probabilities, we estimate the regression given by equation (4), which relates the alternative-specific constants that capture average utilities to vehicle attributes. As noted, we use instrumental variables because price is likely to be correlated with omitted attributes. Nash equilibrium in prices implies that the

[12] Other forms of quasi-random draws have been investigated for use in maximum simulated likelihood estimation of choice models. Sándor and Train (2004) explore (*t,m,s*)-nets, which include Sobol, Faure, Niederreiter and other sequences. They find that Halton draws performed marginally better than two types of nets and marginally worse than two others, and that all the quasi-random methods vastly outperformed pseudo-random draws. In high dimensions, when Halton draws tend to be highly correlated over dimensions, Bhat (2003) has investigated the use of scrambled Halton draws, and Hess et al. (2006) propose modified Latin hypercube sampling procedures. The dimension of integration in our model is not sufficiently high to require these procedures.

price of each vehicle depends on the attributes of all the other vehicles, which indicates that appropriate instruments can be constructed from these attributes because they are unlikely to be correlated with a given vehicle's omitted attributes. Letting d_{ji} be the difference in an attribute, say fuel economy, between vehicle *j* and *i*, we calculate four instruments for vehicle *i* for each attribute: the sum of d_{ji} over all *j* made by the same manufacturer, the sum of d_{ji} over all *j* made by competing manufacturers, the sum of d_{ji}^2 over all *j* made by the same manufacturer, and the sum of d_{ji}^2 over all *j* by competing manufacturers.

The four measures are the instruments obtained from the exchangeable basis developed by Pakes (1994). The first two have been used by Berry, Levinsohn, and Pakes (1995) and Petrin (2002). The latter two measures, which have not been used before, capture the extent to which other vehicles' non-price attributes differ from vehicle *i's* non-price attributes. We found them to be quite useful in our estimations because without them parameter estimates tended to be less stable across alternative specifications.

Estimation of the first stage regressions for price and retained value (the two endogenous variables described further below) obtained R-squares of 0.82 and 0.83 respectively. Based on F-tests, the hypotheses that all instruments have zero coefficients and that the extra instruments that do not also enter as explanatory variables in the second stage have zero coefficients, can be rejected at the 99 percent confidence level. We should point out, however, that use of the instruments assumes that unobserved attributes, while correlated with price, are independent of the observed non-price attributes of vehicles. This assumption, previously maintained by Berry, Levinsohn, and Pakes (1995, 2004) and Petrin (2002), is justified to some extent by pragmatic considerations. In future work, it would be useful to explore the possibility of and remedies to any endogeneity in observed non-price attributes.

4. Model Specification, Data, and Estimation Results

The random utility function in equation *(1)* posits that consumers' vehicle choices and their ranking of vehicles that they seriously considered are determined by vehicle attributes, their socioeconomic characteristics and brand loyalty, and an automaker's product line and distribution network. The regression model specifies the average utility of a given make and model as a function of vehicle attributes.

In addition to a vehicle's purchase price, the attributes that we include in the models are fuel economy, horsepower, curb weight, length, wheelbase, reliability, transmission type, and size classifications. These attributes encompass those used in previous research. Other safety-related variables such as airbags and antilock brakes were not included because most vehicles in our sample were equipped with them. Because automobiles are a capital good, consumers' choices may also be influenced by their expectations of how much a vehicle's value will depreciate. We therefore include as a separate variable the percentage of a vehicle's purchase price in 2000 (consistent with the sample discussed below) that it is expected to retain after two years of ownership. Calculating the retained value based on three years of ownership produced a slightly worse fit than using two years of ownership, while calculating the value based on four years of ownership produced a noticeably worse fit. We expect that consumers are more likely to select a vehicle that retains its value (i.e., the coefficient should have a positive sign) because it could be sold or traded in for a higher price than a vehicle that retains little of its value. As noted, we measure brand loyalty by a consumer's consecutive purchases of the same brand of vehicle. The socioeconomic characteristics that we include are sex, age, income, residential location, and family size.

Our specification extends previous vehicle demand models by exploring the effect of automakers' product line and distribution network on choice. Researchers have typically used brand preference dummy variables to capture these influences. Economic theory suggests that broad product lines can create first mover advantages to a firm and overcome limited information in a market; thus, we specify the number of distinct models (i.e., nameplates) offered by an automaker to capture these possible effects. During the past decade, GM in particular has been criticized for offering too many models that are essentially the same vehicle, suggesting that the sign of this variable may vary by automaker. Industry analysts stress that automakers benefit from having a "hot car" in their product line because it may draw attention to other vehicles that they produce. For many decades, a well-known axiom among the Big Three was: "bring them into the showroom with a convertible, and sell them a station wagon." Recently, GM tried to get buzz for the Pontiac G6 sedan that it hoped would spillover to its other products by giving away 276 of these vehicles on Oprah Winfrey's television show. We constructed a dummy variable that indicated whether a manufacturer produced a hot car, where a hot car was defined as having sales equal to the mean sales of its subclass plus twice the standard deviation of sales. (We also explored other definitions.) An automaker's network of dealers distributes its products to potential customers; thus, we also include the number of each manufacturing division's dealerships.

We performed estimations based on a random sample of 458 consumers who acquired—that is, paid cash, financed, or leased—a new 2000 model year vehicle.[13] Although these consumers differed in how they financed a vehicle, we found that their choice model parameters were not statistically different and thus combined them to estimate a single model. The sample was drawn from a panel of 250,000 nationally representative U.S. households that is aligned with demographic data from the Current Population Survey of the U.S. Bureau of the Census. The panel is administered by National Family Opinion, Inc., and managed by Allison-Fisher, Inc. The response rate for our sample exceeded 70 percent. The data consist of consumers' new vehicle choices by make and model, their ranking of the vehicles they seriously considered acquiring, vehicle ownership histories, which are used to construct the brand loyalty variables, and socioeconomic characteristics. Vehicle attributes and product line variables are from issues of *Consumer Reports,* the *Market Data Book* published by Automotive News, and *Wards' Automotive Yearbook*. We follow previous research and use the manufacturer suggested retail price, MSRP, for the purchase price. Although manufacturers discount these prices with various incentives, such as cash rebates and interest free loans, during our sample period the difference between the incentives offered by American, Japanese, and European manufacturers as a percentage of the retail prices of their vehicles was quite small. Vehicles' expected retained values were obtained from the *Kelley Blue Book*: *Residual Value Guide*. The number of division dealerships within 50 miles of a

[13] The sample size is limited by our requiring data for each consumer on the number of dealers within 50 miles that sell each make/model of vehicle and consumers' vehicle ownership histories and rankings of vehicles they considered in their 2000 choice. This information is not available from standard surveys such as the CES. Our survey enabled us to obtain the information, but at a high cost per respondent.

respondent's zip code was obtained from the automakers' websites. A 50-mile radius seems appropriate for our analysis because CNW Marketing Research found that consumers travel 22 miles, on average, to acquire a new vehicle. In addition, some automakers' web pages only display dealerships within 50 miles of the inputted zip code.

Table 3 provides some descriptive information about the sample. It is difficult to obtain population data to assess the sample because it is conditional on a consumer acquiring a new 2000 model year vehicle. However, as noted, the sample is derived from a panel of U.S. households whose demographics are consistent with national figures; accordingly, the sample values of the socioeconomic characteristics appear to be representative. Moreover, the sample market shares of the manufacturers by geographic origin are well-aligned with the national market shares reported in table 1.

PUT TABLE 3 ABOUT HERE.

Each consumer's choice set consisted of the 200 makes and models of new 2000 vehicles. We treated a number of manufacturers that merged in the late 1990s, for example, Daimler-Benz and Chrysler, as offering distinct makes because it was likely that consumers had not yet perceived that their vehicles were made by the same manufacturer. Indeed, we obtained more satisfactory statistical fits under this assumption than using the merged entity as a unit of analysis. Given this choice set, we estimated a mixed logit model that included brand loyalty, product line and distribution variables, and vehicle attributes interacted with consumer characteristics, error components, and an alternative specific constant for each vehicle make and model. The estimated constants, which capture average utility, were then regressed against vehicle attributes using instrumental variables.

Table 4 presents estimation results for all parts of the model because each part contributes to consumers' utility. The first panel gives coefficients for two specifications of average utility; for reasons explained below, one specification does not include the retained value and the other does. The second panel contains the estimated coefficients for the variation in utility that relates to consumers' observed characteristics; and the third, coefficients for the error components, assumed to be normally distributed, that capture variation in utility that is not related to observed characteristics. Alternative distributions for the error components such as the lognormal did not produce fits as satisfactory as the normal.

PUT TABLE 4 ABOUT HERE.

4.1 Price Coefficients

Consumers' response to a change in the price of a given vehicle is captured by an average effect, an effect that varies with income, and an effect that varies over consumers with the same income. That is, for the model without retained value, the estimate of the derivative of utility with respect to price is:

$$-0.073 - 1.60/\text{consumer income} + 0.86\,\eta/\text{consumer income},$$

where η is distributed standard normal. As previously indicated, the first term is estimated using instrumental variables (IV); when ordinary least squares (OLS) is used the coefficient falls to –0.043 indicating that omitted attributes are correlated with price and that it is important to correct for endogeneity in estimation. Based on these coefficients, the average price elasticity for all vehicles is -2.32, which is consistent with estimates obtained by Berry, Levinsohn, and Pakes (2004).[14]

When a vehicle's expected retained value is specified as an additional explanatory variable, it appears to play an important role in controlling for the endogeneity of price. We isolate this effect in table 5, which reports the coefficients for the purchase price and the retained value estimated by OLS and IV. Given that the retained value is derived from the purchase price, it is likely to be correlated with unobserved attributes of the vehicle and should therefore be estimated by IV. As noted, when we include price but not the retained value in the specification, the difference between the OLS and IV estimates indicated a considerable degree of endogeneity. But when we also include the retained value, it

---

[14] The elasticities are calculated as the percent change in predicted market share that results from a one percent change in price, where predicted market shares are obtained by integrating over both observed and unobserved consumer attributes. A separate elasticity is calculated for each make and model of vehicle. The average given in the text is over all makes and models.

appears to absorb most of the endogeneity bias while the OLS and IV estimates of the purchase price are very similar. This finding suggests that unobserved attributes are correlated with a vehicle's retained value but not with the *difference* between its price and retained value (i.e., expected vehicle depreciation).

PUT TABLE 5 ABOUT HERE.

Note that the retained value represents about 60 percent, on average, of the purchase price (as measured by the MSRP) of a vehicle; thus, the combined effect, regardless of whether it is estimated by OLS or IV, of the retained value and price on average utility is roughly the same as the effect of price when it is entered by itself. This relationship suggests that the model with the retained value effectively decomposes the two components of price to which a consumer responds. Moreover, holding retained value constant, table 5 shows that consumers' response to price (i.e., the average price elasticity) is clearly higher than when the retained value is allowed to vary. The reason is that the retained value is determined by competitive used-vehicle markets; hence, if a manufacturer raises the price of a new vehicle without improving its attributes, the retained value will not rise proportionately and may not rise at all.

As expected, the separate price effects are estimated with less precision than the combined effect. Indeed, the estimated coefficient of retained value obtains a t-statistic of only 0.5, which suggests that the hypothesis that consumers do not differentiate between the two components of price cannot be rejected. Nonetheless, the pattern of estimates is consistent with rational behavior and a plausible form of endogeneity, and may have important implications for estimating the price elasticity that is actually relevant to firms' behavior. It therefore seems reasonable to maintain the concept of retained value as a potential influence among the set of vehicle attributes affecting consumer choice and subject it to further exploration in future research.[15]

---

[15] The inclusion of retained value may alternatively be interpreted as an application of Matzkin's (2004) method of correcting for endogeneity. Retained value would qualify as the extra variable needed for Matzkin's approach if it is related to the price only through

4.2 Other Coefficients

The non-price vehicle attributes in table 4 enter utility with plausible signs and are nearly always statistically significant. Vehicle reliability, horsepower divided by curb weight, automatic transmission included as standard equipment, wheelbase, and vehicle length beyond the wheelbase have a positive effect on the likelihood of choosing a given vehicle, while fuel consumption per mile (the inverse of miles per gallon) has a negative effect. Note that wheelbase tends to reflect the size of the passenger compartment and therefore, as expected, has a larger coefficient than vehicle length beyond the wheelbase. Other measures of vehicle size, such as width and a proxy for interior volume, did not have statistically significant effects. We also performed estimations that included engine size (in liters), but it had a statistically insignificant effect.

Our findings that the (dis)utility of price is inversely related to income and that reliability has a positive and statistically significant effect on utility for women over 30 years of age but has an insignificant effect for men and for women under 30 exemplify observed heterogeneity in consumer preferences. Other examples are that consumers who lease a vehicle are more likely to engage in upgrade behavior by choosing a luxury or sports car than consumers who purchase a vehicle (Mannering, Winston, and Starkey (2002) discuss this phenomenon), and that households with adolescents are more likely than other households to choose a van or SUV presumably to use for work and non-work trips.

---

exogenous perturbations, but is correlated with the unobserved attributes of a vehicle. Under these conditions, the original error term may be expressed as a function of the retained value and a new error term that is independent of all explanatory variables including price, which would permit OLS estimation of the regression to yield consistent parameter estimates. As expected from an endogeneity correction, the OLS estimate of the price coefficient rises when the retained value is included in the model (compare the OLS estimate in the third column of table 5 with the OLS estimate in the first column) and is similar to the IV estimate of the price coefficient (in the second column). We also estimated the function of retained value non-parametrically and obtained essentially the same results as when we specified retained value linearly.

Unobserved preference heterogeneity is captured in error components related to vehicle price, horsepower, fuel consumption, and consumers' preferences for cars versus trucks (including light trucks, vans, and SUVs).[16] The last coefficient reflects greater substitution among cars and among trucks than across these categories, which is confirmed by our estimates of vehicle cross-elasticities. For example, we find that the cross-elasticity of demand with respect to the price of a given make and model of a van is, on average: 0.038 for other makes and models of vans; 0.026 for makes and models of SUVs; 0.018 for makes and models of pickup trucks; 0.0025 for makes and models of regular cars; and 0.0021 for makes and models of sports and luxury vehicles.[17] As expected, cross-elasticities are higher for more similar types of vehicles. We also found reasonable cross-elasticity patterns for the prices of other vehicle types. In contrast, a model that maintained the IIA property would restrict the cross-elasticity of demand with respect to a given vehicle's price to be the same for all vehicles; that is, IIA implies that the elasticity of vehicle j's demand with respect to a change in vehicle i's price is the same for all $j \neq i$. .

Surprisingly, we found that, all else constant, consumers were not more likely to purchase a vehicle from automakers that offered a large (or small) number of models or that produced a "hot car." We explored various definitions of a hot car to construct its dummy variable, based on deviations from mean sales and sales growth, but they were all statistically insignificant. We also specified hot car dummies based on vehicle size

[16] These components were determined after extensive testing of a variety of specifications, including models that allowed the densities to depend on income and other variables. We were not able to identify any other statistically significant influences on the components beyond those captured in the fixed portion of utility (i.e., the mean of the error components). Recall that we could not identify significant error components without including data on considered vehicles, which suggests that the data contain limited information on the distribution of unobserved taste variation.

[17] To put the magnitude of the cross-elasticities in perspective: if a vehicle had a market share of 0.005 (i.e., the average share because there are 200 makes and models of vehicles) and had an own-price elasticity of –3.0, then the cross-price elasticity for each other vehicle, assuming it did not vary, would be 0.0151.

classifications but they were also statistically insignificant. Although automakers cannot rely on product line "externalities" to improve their sales, we found that their dealer network does have a statistically significant effect on choice. We constructed the dealership variable by division as the natural log of one plus the actual number of dealers within 50 miles of the consumer up to a maximum of three. Thus, the variable takes on a value of zero if no dealers within the circumscribed area sell the vehicle. In addition, the functional form assumes that the impact of having one dealer instead of none is greater than the extra impact of having a second dealer instead of one, and so on, with the impact of additional dealers negligible beyond three. This specification fit the data better than a linear specification, indicating that it is important for automakers to have a dealer within reasonable proximity to potential customers but that additional dealers will have a diminishing impact on sales.

Finally, we included separate brand loyalty variables for GM, Ford, and Chrysler as well as for the Japanese and European automakers as distinct groups. Preliminary estimations indicated that it was statistically justifiable to aggregate the Japanese and European automakers into single loyalty variables. We could not estimate a brand loyalty parameter for Korean automakers because only one consumer in the sample chose a Korean vehicle in their most recent previous purchase. The estimated coefficients are positive, statistically significant, and fairly large while the error component for brand (manufacturer) loyalty is statistically significant. We found that the likelihood function increased when we used the conditional distribution of η_n rather than its unconditional distribution, which indicates that conditioning provides useful information about consumers' choices.

When our estimates are assessed in the context of previous findings that use the same measure of brand loyalty as used here, it becomes clear that loyalties have undergone considerable shifts as consumers have gained experience with and adjusted to new information about automakers' products. Mannering and Winston (1991) found that during the 1970s, American consumers had the greatest brand loyalty toward Chrysler, had comparable loyalty toward GM and Japanese automakers, and the least loyalty for Ford. During the 1980s, after American consumers developed greater experience with Japanese vehicles, Mannering and Winston found that loyalty toward Japanese automakers exceeded loyalty toward any American automaker. But during the mid-1990s, as American consumers gained experience with certain automakers by leasing their vehicles and purchasing a greater share of light trucks, Mannering, Winston, and Starkey (2002) found that American consumers developed strong brand loyalty toward European automakers and revived some of their loyalty toward American firms.

Our brand loyalty estimates indicate that this recent shift is intact because consumers have the strongest loyalty toward European automakers while loyalty for Ford and Chrysler now exceeds loyalty toward Japanese automakers. Of course, Ford's and Chrysler's loyalty coefficients may indicate that as their market shares have fallen, they have retained a smaller but more loyal group of customers. GM has the least loyalty and, in contrast to Ford and Chrysler, appears to be retaining only loyal rural customers as its share falls.

We stress that our interpretations should be qualified on theoretical grounds because the loyalty coefficients could also be capturing heterogeneity in tastes. We cannot resolve the theoretical debate, but we did explore additional empirical treatments of brand loyalty to shed light on the validity of our interpretation. In particular, if the phenomenon we are capturing were unobserved tastes for vehicle types, then it is likely that such tastes would be correlated with at least some of the vehicle attributes in the model. But, as noted earlier, when we performed estimations without a manufacturer error component and without including the brand loyalty variables, the other (non-brand loyalty) parameters were nearly the same as those presented in table 4. Of course, this exploration does not rule out the possibility that the loyalty variables themselves are subject to endogeneity bias; but at a minimum it indicates that such bias does not affect the other parameters of the model, which is an important consideration when we assess the sources of changes in market shares.

5. Assessing the U.S. Automakers' Decline

The main purpose of the vehicle choice model is to guide an exploratory assessment of the ongoing decline in U.S. automakers' market share. As discussed in the introduction, several hypotheses that explain the decline could be derived from the academic literature and the views of industry observers and participants including changes in basic vehicle attributes, subtle vehicle attributes, unobserved tastes, brand loyalty, product line characteristics, and distribution outlets.

The findings obtained from the vehicle choice model narrow the range of possible explanations to vehicle attributes and unobserved tastes. The statistically insignificant parameter estimates for the product line variables and the apparent relative improvement in brand loyalty for Ford and Chrysler suggest that these factors are unlikely to have been a major source of the industry's loss in market share. Foreign automakers have improved their distribution networks over time, but U.S. automakers compete effectively in this dimension. Thus, we first focus on the impact of changes in basic vehicle attributes during the past decade on U.S. automakers' market shares and if necessary turn to less observable factors.

We use data on the vehicles offered in 1990 and their attributes to forecast the change in U.S. automakers' market share attributable to changes in vehicle attributes given consumers' tastes in 2000. Data for vehicle offerings and attributes in 1990 were obtained from *Consumer Reports*, Automotive News' *Market Data Book*, and *Wards' Automotive Yearbook*. Prices for vehicles in 1990 were expressed in 2000 dollars. By construction, forecasted shares equal actual shares in 2000 when the forecasts are obtained with the choice probabilities P_{ni} estimated in table 4. These forecasts rely on δ_j for all j, including its unobserved component ξ_j. The values of the ξ_j 's are not known for vehicles in any year other than that used in estimation. To forecast what market shares would have been in 2000 given 1990 basic vehicle attributes and offerings, we adopted an approach that is similar to that implemented by Berry, Levinsohn, and Pakes (2004). For any 1990 vehicle that was still offered under the same model name in 2000, we used the estimated value of ξ_j for that vehicle in 2000. For 1990 vehicles that did not continue into 2000, we used the average of ξ_j over 2000 vehicles of the same type (i.e., SUV, van, pickup, sports, and other) by the same manufacturer (with Japanese, European, and other manufacturers each grouped.)[18]

[18] We obtained an indication of the impact of this type of averaging of the ξ_j 's by applying the procedure in forecasts for 2000, using the estimated ξ_j for 2000 vehicles that also existed in 1990 and using the manufacturer/type averages for 2000 vehicles that did not also exist in 1990. The forecasted share of US manufactures based on this procedure was

By utilizing this procedure for the ξ_j 's, our forecasts (and changes in shares) represent the impact of changes in the observed basic attributes of vehicles between 1990 and 2000 but not changes in unobserved attributes. As noted below, we explored two other procedures for treating the ξ_j 's in our forecasts.

Market shares are forecasted for the 1990 vehicle offerings and attributes, thereby allowing us to compare consumers' 2000 choices with a prediction of what vehicles they would have purchased in 2000 had they been offered the vehicles (and attributes) that were available in 1990. A simple consumer surplus calculation based on the familiar "log sum" expression for the logit model indicated that all of the automakers (by geographical origin) improved the attributes of their vehicles over the decade. Thus, the change in U.S. automakers' market share predicted by the model reflects the *relative* improvement in their vehicles.

We find that the relative change in American manufacturers' offerings and attributes was responsible for the industry losing 6.34 percentage points of market share, which accounts for almost all of the 6.80 percentage points of market share that the U.S. industry actually lost during the past decade.[19] Our sample is not large enough to provide reliable breakdowns by automaker and vehicle classification; however, we can

---

0.65625 compared with the actual share of 0.65650, indicating that averaging has little impact on forecasts of US manufacturers' share.

[19] We also forecasted the changes in market shares using two other ways of handling the unobserved attributes of vehicles, ξ_j. In one procedure, we integrated the choice probabilities over the empirical distribution of the unobserved attributes. That is, for each vehicle we randomly chose a value of ξ_j from the values estimated for the year 2000 vehicles; we repeated the forecasts numerous times and averaged the results. The estimated change in market share for U.S. manufacturers was 6.71, which is even closer to the 6.80 change that actually occurred. Second, following a suggestion of Steven Berry, we used a variant on this integration procedure in which the correlation between price and unobserved attributes is incorporated. The estimated change was essentially the same as in the first procedure.

report that virtually all segments of the American manufacturers' products experienced some loss in market share. This important but disturbing finding suggests that although the American industry has received various kinds of trade protection for more than two decades ostensibly to help it "retool" and has benefited from robust macroeconomic expansions during the 1980s and 1990s, it continues to lag behind foreign competitors when it comes to producing a vehicle with desirable attributes. It is particularly noteworthy that the loss of the American industry's market share can be explained by changes in the basic attributes—price, fuel consumption, horsepower, and so on—that are included in our model, rather than subtle attributes such as styling and various options or unobserved tastes.[20]

We performed a simulation to determine how much U.S. manufacturers would have to reduce their prices in 2000 to attain the same market share in 2000 that they had in 1990 and found that prices would have to fall more than 50 percent. This large price reduction is reasonable because U.S. manufacturers' market share in 2000 is roughly two-thirds and the price elasticity with respect to a simultaneous change in all U.S. vehicle prices is small. (The price elasticities between -2.0 and -3.0 that we reported previously refer to the change in the price of an *individual* make and model of a vehicle.) Although it would not be profit maximizing for U.S. firms to contemplate such a strategy, they have recently attempted to retain and possibly recover some of their market share by offering much larger incentives than foreign automakers offer. However, even this short-term fix has had little effect on their sales; as suggested by our simulation, the price reductions that would be needed to affect their share are considerably larger than those that have been offered. Indeed, despite offering

---

[20] We also forecast the impact of the changes in dealership networks that occurred from 1990 to 2000 and found that the change in dealership networks resulted in a loss of 0.5 percentage points for U.S. manufacturers. This predicted loss is very small, indicating that the relative improvement in foreign automakers' networks is not an important factor in the decline of U.S. manufacturers' share. However, combining this loss in share with the loss due to changes in basic vehicle attributes enables us to account for the entire loss of 6.8 percentage points that actually occurred.

incentives in 2005 that were as much as $3,000 per vehicle greater than the incentives offered by Japanese manufacturers, U.S. automakers' market share of cars and light trucks in that year fell 2 percentage points from its share in the previous year.

In contrast to the U.S. automakers, European firms' market share increased some five percentage points over the decade, partly because they intensified competitive pressure on the U.S. automakers by offering attractive entry-level luxury vehicles such as the restyled BMW 3-series. Indeed, European automakers achieved a net gain of 12 new vehicle models over the decade, while U.S. and Japanese automakers' net change was negligible. Japanese automakers gained roughly a percentage point of share as they expanded their presence in the higher (and more profitable) end of the market with various new offerings from Lexus, Infiniti, and Acura.

6. Conclusion

Concerns about the competitiveness of the U.S. automobile industry developed in the early 1980s when Chrysler needed a bailout from the federal government to avoid financial collapse and Ford and General Motors suffered large losses. Since then, the profitability of the domestic industry has fluctuated while its market share has steadily declined. Investors in the stock market, who are the most experienced and credible soothsayers of an industry's future, envision that difficult times lie ahead for Ford, General Motors, and Daimler-Chrysler as the sum of their current market capitalization is less than half the combined market capitalization of Honda, Toyota, and Nissan and less than Toyota's market capitalization alone. Toyota's consistent profitability has allowed it to invest in fuel-efficient hybrid engine systems for compact and luxury cars, and to take risks, like starting a youth-focused brand, Scion, thereby increasing pressure on other automakers.

We have attempted to shed light on the U.S. industry's current predicament by applying recent econometric advances to analyze the vehicle choices of American consumers. Notwithstanding these advances, we have been confronted with some formidable methodological challenges that necessitated some compromises. We have identified the advantages and limitations of our approach while setting the stage for future research.

We have found that the U.S. automakers' loss in market share during the past decade can be explained almost entirely by the difference in the basic attributes that measure the quality and value of their vehicles. Recent efforts by U.S. firms to offset this disadvantage by offering much larger incentives than foreign automakers offer have not met with much success. In contrast to the numerous hypotheses that have been proffered to explain the industry's problems, our findings lead to the conclusion that the only way for the U.S. industry to stop its decline is to improve the basic attributes of their vehicles as rapidly as foreign competitors have been able to improve the basic attributes of theirs. The failure of U.S. automobile firms to address this fundamental deficiency suggests that these organizations may be saddled with constraints that researchers and industry analysts have yet to identify.

Authors' affliations:
Kenneth E. Train, Department of Economics, University of California, Berkeley, U.S.
Clifford Winston, Brookings Institution, U.S.

References

Aizcorbe, A., C. Winston, and A. Friedlaender, "Cost Competitiveness of the U.S. Automobile Industry," in Clifford Winston and Associates, Blind Intersection? Policy and the Automobile Industry (Washington, DC: Brookings, 1987).

Berry, S., "Estimating Discrete-Choice Models of Product Differentiation," RAND Journal of Economics 25 (1994), 242-262.

Berry, S., J. Levinsohn, and A. Pakes, "Automobile Prices in Equilibrium," Econometrica 63 (1995), 841-890.

Berry, S., J. Levinsohn, and A. Pakes, "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Vehicle Market," Journal of Political Economy 112 (2004), 68-105.

Bhat, C., "Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," Transportation Research Part B 35 (2001), 677-693.

Bhat, C., "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences," Transportation Research Part B 37 (2003), 837-855.

Boyd, J. and R. Mellman, "The Effect of Fuel Economy Standards on the U.S. Automotive Market: A Hedonic Demand Analysis," Transportation Research Part A 14 (1980), 367-378.

Brownstone, D. and K. Train, "Forecasting New Product Penetration with Flexible Substitution Patterns," Journal of Econometrics 89 (1999), 109-129.

Cardell, S., N. Scott, and F. Dunbar, Measuring the Societal Impacts of Automobile Downsizing," Transportation Research Part A 14 (1980), 423-434.

Goolsbee, A., and A. Petrin, "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV," Econometrica 72 (2004), 351-381.

Heckman, J., "Statistical Models for the Analysis of Discrete Panel Data," in C. Manski and D. McFadden, editors, Structural Analysis of Discrete Data (Cambridge: MIT Press, 1981a).

Heckman, J., "The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Stochastic Process and Some Monte Carlo Evidence on Their Practical Importance," in C. Manski and D. McFadden, editors, Structural Analysis of Discrete Data (Cambridge: MIT Press, 1981b).

Heckman, J., "Identifying the Hand of the Past: Distinguishing State Dependence from Heterogeneity," American Economic Review 81 (1991), 75-79.

Hess, S., K. Train, and J. Polak, "On the Use of a Modified Latin Hypercube Sampling (MLHS) Method in the Estimation of a Mixed Logit Model for Vehicle Choice," forthcoming, Transportation Research Part B 40 (2006), 147-167.

Lave, C., and K. Train, "A Disaggregate Model of Auto-Type Choice," Transportation Research Part A 13 (1979), 1-9.

Luce, R. and P. Suppes, "Preference, Utility, and Subjective Probability," in R. Luce, R. Bush, and E. Galanter, editors, Handbook of Mathematical Psychology, III (New York: Wiley, 1965).

Mannering, F. and C. Winston , "A Dynamic Empirical Analysis of Household Vehicle Ownership and Utilization," RAND Journal of Economics 16 (1985), 215-236.

Mannering, F.. and C. Winston, "Brand Loyalty and the Decline of American Automobile Firms," Brookings Papers on Economic Activity: Microeconomics 1991, 67-114.

Mannering, F., C. Winston, and W. Starkey, “An Exploratory Analysis of Automobile Leasing by U.S. Households,” Journal of Urban Economics 52 (2002),154-176.

Manski, C., and L. Sherman, “An Empirical Analysis of Household Choice Among Motor Vehicles,” Transportation Research Part A 14 (1980), 349-366.

Matzkin, R., “Unobservable Instruments,” working paper, Department of Economics, Northwestern University, 2004.

McFadden, D., and K. Train, “Mixed MNL Models for Discrete Response,” Journal of Applied Econometrics 15 (2000), 447-470.

Pakes, A., “Dynamic Structural Models, Problems and Prospects: Mixed Continuous/Discrete Controls and Market Interactions,” in J-J. Laffont and C. Sims, editors, Advances in Econometrics: The Sixth World Congress of the Econometric Society, Vol. II (New York: Cambridge University Press, 1994).

Petrin, A., “Quantifying the Benefits of New Products: The Case of the Minivan,” Journal of Political Economy 110 (2002), 705-729.

Revelt, D. and K. Train, “Mixed Logit with Repeated Choices: Households’ Choice of Appliance Efficiency Level,” Review of Economics and Statistics 80 (1998), 647-657.

Sándor, Z. and K. Train , “Quasi-random Simulation of Discrete Choice Models,” Transportation Research Part B 38 (2004), 313-327.

Train, K., Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand (Cambridge: MIT Press, 1986).

Train, K., "Halton Sequences for Mixed Logit," working paper, Department of Economics, University of California, Berkeley, 2000.

Table 1. U.S. and Foreign Automakers' Market Share of Vehicle Sales in the U.S.

		Manufacturer by Geographic Origin		
	Year	US	Japan	Europe
Market Share of Cars (percent)	1970	86	3	8
	1975	82	9	7
	1980	74	20	6
	1985	75	20	5
	1990	67	30	5
	1995	61	31	5
	2000	53	32	11
	2005	42	40	11
Market Share of Light Trucks (percent)**	1970	91	4	4
	1975	93	6	1
	1980	87	11	2
	1985	81	18	0
	1990	84	16	0
	1995	87	13	0
	2000	77	19	1
	2005	70	25	3
Market Share of Cars and Light Trucks (percent)	1970	87	4	7
	1975	85	8	6
	1980	77	18	6
	1985	77	19	4
	1990	72	24	3
	1995	72	23	3
	2000	66	26	6
	2005	57	32	7

* Shares generally do not sum to 100 because of rounding, the omission of Korean manufacturers, and imports that Automotive News does not assign to any manufacturer or country of origin.

**Light trucks include SUVs, minivans, and pickups weighing over 6000 pounds.

Source: Automotive News Market Data Book (1980-2006)

Table 2. “Big Three” and Selected Foreign Automakers’ Market Share of Vehicle Sales in the U.S.

		Manufacturer				
	Year	General Motors	Ford	Chrysler (Domestic)	Toyota	Honda
Market Share of Cars (percent)	1970	40	26	16	2	0
	1975	44	23	11	3	1
	1980	46	17	9	6	4
	1985	43	19	11	5	5
	1990	36	21	9	8	9
	1995	31	21	9	9	9
	2000	28	17	8	11	10
	2005	22	13	9	16	11
Market Share of Light Trucks (percent)*	1970	38	38	9	1	0
	1975	42	31	15	2	0
	1980	39	33	11	6	0
	1985	36	27	14	7	0
	1990	35	30	14	6	0
	1995	31	33	16	5	1
	2000	28	28	15	8	3
	2005	30	23	18	11	6
Market Share of Cars and Light Trucks (percent)						
	1970	40	28	15	2	0
	1975	43	25	12	3	1
	1980	45	20	9	6	3
	1985	41	21	12	6	4
	1990	35	24	11	8	6
	1995	31	26	12	7	5
	2000	28	23	12	9	7
	2005	26	19	14	13	9

*Light trucks include SUVs, minivans, and pickups weighing over 6000 pounds. AMC/Jeep was acquired by Chrysler in 1987, but is not included in Chrysler’s share to maintain consistency over time. Source: Automotive News Market Data Book (1980-2006)

Table 3. Description of the Sample
(Consumers who acquired a new vehicle in the year 2000)

Socioeconomic Characteristics	
Variable	Sample Value
Average Household Income	$67,767
Average Age	54.2
Percent male	54
Percent with child aged 1-16	19
Percent who live in rural location [a]	45

Market Share of Cars and Light Trucks by Manufacturer's Geographic Origin:	
Manufacturer	Share (percent)
U.S.	64
Japanese	28
European	5
Other	3

[a] A rural location is defined as being outside of an MSA of 1 million people or more.

Table 4. Vehicle Demand Model Parameter Estimates*

Average utility: elements of $\alpha' z_j$	Coefficient (Standard Error)	
Constant	-7.0318 (1.4884)	-6.8520 (1.5274)
Manufacturer's suggested retail price *(in thousands of 2000 dollars)*	-0.0733 (0.0192)	-0.1063 (0.0635)
Expected retained value after 2 years *(in thousands of 2000 dollars)*	---	0.0550 (0.1011)
Horsepower divided by weight *(in tons)*	0.0328 (0.0117)	0.0312 (0.0120)
Automatic transmission dummy *(1 if automatic transmission is standard equipment; 0 otherwise)*	0.6523 (0.2807)	0.6787 (0.2853)
Wheelbase *(inches)*	0.0516 (0.0127)	0.0509 (0.0128)
Length minus wheelbase *(inches)*	0.0278 (0.0069)	0.0279 (0.0069)
Fuel consumption *(in gallons per mile)*	-0.0032 (0.0023)	-0.0032 (0.0023)
Luxury or sports car dummy *(1 if vehicle is a luxury or sports car, 0 otherwise)*	-0.0686 (0.2711)	-0.0558 (0.2726)
SUV or station wagon dummy *(1 if vehicle is a SUV or wagon, 0 otherwise)*	0.7535 (0.4253)	0.7231 (0.4298)
Minivan and full-sized van dummy *(1 if vehicle is a minivan or full-sized van, 0 otherwise)*	-1.1230 (0.3748)	-1.1288 (0.3757)
Pickup truck dummy *(1 if the vehicle is a pickup truck, 0 otherwise)*	0.0747 (0.4745)	0.0661 (0.4756)

Chrysler manufacturer dummy	0.0228 (0.2794)	0.0654 (0.2906)
Ford manufacturer dummy	0.1941 (0.2808)	0.2696 (0.3060)
General Motors manufacturer dummy	0.3169 (0.2292)	0.3715 (0.2507)
European manufacturer dummy	2.4643 (0.3424)	2.4008 (0.3624)
Korean manufacturer dummy	0.7340 (0.3910)	0.8017 (0.4111)

Utility that varies over consumers related to observed characteristics: elements of $\beta' x_{nj}$	Coefficient (Standard Error)
Manufacturers' suggested retail price divided by respondent's income	-1.6025 (0.4260)
Vehicle reliability based on the *Consumer Reports'* repair index for women aged 30 or over *(0 otherwise)* [a]	0.3949 (0.0588)
Luxury or sports car dummy for lessors *(1 if the vehicle is a luxury or sports car and the respondent leased , 0 otherwise)*	0.6778 (0.4803)
Minivan and full-sized van dummy for households with an adolescent *(1 if the vehicle is a van and the respondent's household has children aged 7 to 16, 0 otherwise)*	3.2337 (0.5018)
SUV or station wagon dummy for households with an adolescent *(1 if vehicle is a SUV or Wagon and the respondent's household includes a child aged 7 to 16, 0 otherwise)*	2.0420 (0.4765)
ln(1+Number of dealerships within 50 Miles of the center of a respondent's zip code) [b]	1.4307 (0.2714)
Number of previous consecutive GM purchases	0.3724 (0.1471)
Number of previous consecutive GM purchases for respondents who live in a rural location [c]	0.3304 (0.2221)
Number of previous consecutive Ford purchases	1.1822 (0.1498)

Number of previous consecutive Chrysler purchases	0.9652 (0.2010)
Number of previous consecutive Japanese manufacturer purchases	0.7560 (0.2255)
Number of previous consecutive European manufacturer purchases	1.7252 (0.4657)
Utility that varies over consumers unrelated to observed characteristics (error components): elements of $\mu_n' w_{nj} + \lambda \eta_{ns}$	Coefficient (Standard Error)
Manufacturer's suggested retail price divided by respondent's income times a random standard normal	0.8602 (0.4143)
Horsepower times a random standard normal	45.06 (72.34)
Fuel consumption *(gallons of gasoline per mile)* times a random standard normal	-0.0102 (0.0020)
Light truck, van, or pickup dummy *(1 if vehicle is a light truck, van, or pickup truck; 0 otherwise)* times a random standard normal	6.8505 (2.5572)
Manufacturer loyalty: conditional standard normal as described in text.	0.3453 (0.1712)

*Estimated coefficients for vehicle make and model dummies not shown.

Number of observations: 458

Log likelihood at convergence for choice model: -1994.93

R^2 for regression model: 0.394 without retained value, 0.395 with retained value.

a. The *Consumer Reports'* repair index is a measure of reliability that uses integer values from 1 to 5. A measure of 1 indicates the vehicle has a "much below average" repair record, 3 is "average," while 5 represents "much better than average" reliability.

b. A dealership is defined as a retail location capable of selling a vehicle produced by a given division. The dealership variable is equal to 0,1,2, or 3 (with 3 representing areas with 3 or more dealerships within a fifty-mile radius of the center of the respondent's zip code). This variable is defined for divisions (not manufacturers), because a

Chevrolet dealership might sell Chevrolet vehicles without selling Saturn vehicles (GM manufactures both Chevrolet and Saturn).

c. A respondent is classified as living in a rural location if he or she does not live in a Metropolitan Statistical Area or lives in a Metropolitan Statistical Area with less than 1 million people.

Table 5. Estimated Price Coefficients and Elasticities for Models With and Without the Retained Value

	Model without retained value		Model with retained value	
	OLS	IV	OLS	IV
Purchase Price	-0.043 (0.0094)	-0.073 (0.0192)	-0.122 (0.0362)	-0.106 (0.0635)
Retained value	---	---	0.130 (0.0577)	0.055 (0.1011)
Average price elasticity	-1.7	-2.3	-3.2	-2.9