EE
7/28/06
2:15 PM

Draft – Comments Invited

Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes

Antonio M. Bento
University of Maryland

Lawrence H. Goulder
Stanford University, Resources for the Future, and NBER

Mark R. Jacobsen
Stanford University

Roger H. von Haefen
North Carolina State University

July 2006

We are grateful to David Austin, Tim Bresnahan, Don Fullerton, Michael Greenstone, Ken Small, James Sweeney, Kenneth Train, Kurt van Dender and Frank Wolak for very helpful suggestions, and to Emeric Henry for his important contributions to the econometric efforts. We also thank the William and Flora Hewlett Foundation for financial support.

I. Introduction

For several reasons, reducing automobile-based gasoline consumption is a major U.S. public policy issue. Gasoline consumption accounts for 44 percent of the U.S. demand for crude oil, and the nation's dependence on imported crude raises significant national security concerns. National security impacts can be regarded as an externality from gasoline consumption that gives a potential rationale for public policy. Gasoline use generates environmental externalities as well. The combustion of gasoline yields emissions of carbon dioxide (CO_2), the principal anthropogenically sourced "greenhouse gas" that can contribute to global climate change. In 2004, approximately 22 percent of U.S. CO_2 emissions derived from gasoline use. Gasoline combustion also generates emissions of several "local" air pollutants, including carbon monoxide, nitrogen oxides, and volatile organic compounds. Reduced gasoline use could lead to improved air quality and associated benefits to health.[1,2]

Recently, many analysts have called for new or more stringent policies to discourage gasoline consumption. Proposals include a tightening of corporate average fuel economy (CAFE) standards, subsidies to retirements of older (gas-guzzling) vehicles, subsidies toward the purchase of high-mileage vehicles such as hybrid cars, and increments to the federal gasoline tax.[3,4]

---

[1] Parry and Small (2005) and the National Research Council (2002) examine the various externalities from gasoline use and offer estimates of the overall marginal damages. The former study estimates the overall external cost from U.S. gasoline consumption (including effects relating to local pollution, climate change, congestion, and accidents) to be about 75 cents per gallon. This suggests that U.S. taxes on gasoline are below the efficiency-maximizing level, since the federal tax plus average state tax totals 37 cents.

[2] The extent of the health improvement from improved air quality depends on both the reduction in gasoline use and possible changes in pollution per gallon of gasoline used. Air districts currently in compliance with air pollution regulations under the 1990 Clean Air Act amendments might well respond to reductions in gasoline use by relaxing "tailpipe" emissions requirements, that is, on the allowable emissions per unit of fuel combusted. This would offset the air-quality and health improvements from reduced gasoline consumption.

[3] The National Research Council (2002) has recommended increases in corporate average fuel economy (CAFE) standards on both passenger cars and light trucks. AB 1493, a recently passed bill in California, mandates carbon dioxide emissions reductions and significant improvements in automobile fuel economy. .

[4] Although there is considerable political opposition to many of these proposals, the general public appears to have grown increasingly sympathetic to the need for a significant policy response. A February 2006 New York Times / CBS News Poll found that a majority of Americans would support a higher gasoline tax if it reduced global warming or made the U.S. less dependent on foreign oil.

This paper examines the gas-tax option, employing an econometrically-based multi-market simulation model to evaluate the policy's efficiency and distributional implications. We investigate the economy-wide costs of gasoline tax increases and the impacts on fuel consumption, and we diagnose the sources of the reductions in fuel consumption in terms of changes in fleet composition (shifts to higher mileage automobiles) and vehicle miles traveled (VMT). We also explore how the costs are distributed across households that differ by income, region of residence, race, and other characteristics, and consider how the distribution of impacts depends on the ways revenues from the tax are returned to the private sector.

Some prior studies have examined the impact of gasoline taxes by estimating the demand for gasoline as a function of gasoline price and household income. For example, Hausman and Newey (1995) and West and Williams (2005) use household-level data on gasoline consumption to estimate deadweight loss and the optimal gasoline tax, respectively.

Other studies obtain the demand for gasoline indirectly by deriving it from estimates of the demand for VMT. For example, Berkovec (1985), Mannering and Winston (1985), Train (1986), and West (2004) sequentially estimate households' automobile ownership and VMT decisions.[5] The ownership decision is based on discrete choice of a range of automobile types, while the VMT decision is a continuous choice. Feng, Fullerton, and Gan (2005) estimate jointly households' ownership and VMT decisions.

A third set of studies considers supply-side phenomena that influence the composition of the automobile fleet and gasoline consumption. In contrast with the previously mentioned studies, which tend to focus on issues in public economics (e.g., the deadweight loss from gasoline taxation), this third set of studies emphasizes market structure (e.g., the imperfectly competitive nature of the new car market) and the impacts of policies on new car prices and sales. For example, Berry *et al.* (1995), Goldberg (1998) and Austin and Dinan (2005) develop models of new car market that combine supply decisions by imperfectly competitive producers with discrete demand choices by households. The latter two studies explore impacts of automobile policies on the new car market. Goldberg (1998) analyzes tighter

[5] Poterba (1991) generates elasticities of demand for gasoline use by transforming Train's (1986) estimated elasticities for VMT.

CAFE standards; Austin and Dinan (2005) examine CAFE standards and a gasoline tax increase.

The present study differs from earlier work in several ways. First, in contrast with nearly all prior work,[6] this study considers supply and equilibrium not only in the new car market but in the used car and scrap markets as well. The wider scope helps provide a more complete picture of the impact of a gasoline tax. In addition, addressing the equilibrium in all three car markets enables us to capture important dynamic effects. Higher gasoline taxes are likely to cause an increase in the share of relatively fuel-efficient cars among new cars sold. The extent to which the fuel-efficiency of the *overall* (new and used car) fleet improves will depend on the rate at which the newer, more efficient cars replace older cars. This depends on the relative size of the stocks of new and used cars and the rate at which older cars are taken out of operation (scrapped). By considering the new, used, and scrapped car markets, the model is able to consider the dynamics of changes in fleet composition and related short- and long-run impacts on gasoline consumption. As in Goldberg (1995), Berry, Levinsohn, and Pakes (1995), Petrin (2003), and Austin and Dinan (2005), we consider the imperfectly competitive nature of the new car market. However, in contrast with these studies, we connect this market to the used and scrap markets, which allows us to consider how policies affect the entire fleet of cars and associated demands for gasoline.

A second major difference from earlier work is the model's ability to capture distributional effects. The model considers over 20,000 households that differ in terms of income, family size, employment status (working or retired), region of residence, and ethnic background. This enables us to trace distributional impacts in several important dimensions. All household demands stem from a consistent, utility maximization framework, enabling us to measure distributional impacts in terms of theoretically sound welfare indexes. In contrast with prior simulation studies, the model considers both the acquisition and disposition (recycling) of government revenue, which allows us to examine how distributional impacts are influenced by the way gasoline tax revenues are recycled to the private sector.

Finally, the model differs in its econometric approach to estimating consumer demand for automobiles and gasoline/VMT. Some prior studies have focused separately on the

---

[6] One exception is Berkovec (1985), who develops a model with interactions among these markets. His model assumes pure competition among auto producers, however.

demand for automobiles versus the demand for gasoline or VMT.[7] Other earlier studies have accounted for the simultaneity of automobile and gasoline/VMT demand by exploiting reduced-form, two-step estimators.[8] We adopt a one-step (i.e., full-information) estimation strategy that integrates each household's automobile and VMT decisions within a utility-theoretic framework and permits us to recover sound welfare estimates.[9] In addition, we assume that all parameters entering preferences vary randomly across households. Random coefficients allow us to account for correlations in the unobservable factors influencing a household's discrete car choice and continuous VMT demand while simultaneously allowing for more plausible substitution patterns among automobiles (McFadden and Train, 2000; Bunch *et al.*, 2000).

The rest of the paper is organized as follows. Section 2 describes the equilibrium simulation model. Section 3 outlines the model's data sources, with emphasis on the data employed to estimate household demands for vehicles and travel. Section 4 presents our approach for estimating households' automobile purchase and driving decisions. In Section 5 we present and interpret results from simulations of a range of gasoline tax policies. Section 6 offers conclusions.

2. Structure of the Simulation Model

a. Overview

The economic agents in the model are households, producers of new cars, used car suppliers, and scrap firms. The model considers the car-ownership and vehicle-miles-

[7] Berry, Levinsohn, and Pakes (2004), Goldberg (1995), and Petrin (2003) focus on the demand for automobiles; Goldberg (1998) and West (2004) estimate the demand for VMT; Hausman and Newey (1995), Schmalensee and Stoker (1999), and West and Williams (2005) concentrate on the demand for gasoline. Austin and Dinan (2005) obtain demand functions for cars by calibrating the parameters of their simulation model to be consistent with internal estimates by General Motors.

[8] Examples are Berkovec (1986), Mannering and Winston (1986), Goldberg (1998), and West (2004).

[9] A difficulty with welfare measurement from two-step estimators is that each step generates a different set of estimates for the same parameters. Each set may have different welfare implications for the same policy. One-step estimators generate a unique set of parameter estimates and therefore avoid this difficulty. The only other automobile study to incorporate a one-step procedure is that of Feng, Fullerton, and Gan (2005).

traveled (VMT) decisions of 20,429 households. The car-ownership and VMT decisions are made simultaneously in accordance with utility maximization.

The model distinguishes cars according to age, class, and manufacturer. Table 2-1 below displays the different car categories in the model. These categories imply 350 distinct cars, of which 284 appear in our dataset and simulation.[10]

The used-car market equates the supply of used cars (after scrapping) with the demand for those cars. Producers of new cars decide on new-car prices in accordance with Bertrand (price) competition. These producers consider households' demand functions in determining optimal pricing. Price-markups reflect the various price elasticities of demand for cars as well as existing regulations (CAFE standards).

The model solves for a sequence of market equilibria at one-year intervals. Car vintages are updated each year, so that last year's new cars become one-year-old cars, last year's one-year-old cars become two-year-old cars, etc. Once a car is scrapped, it cannot re-enter the used car market. Characteristics of given models of new cars change through time based on specified growth rates of horsepower and fuel-economy, as described in Section 5.

b. Household Demands

Households obtain utility from car ownership and use, as well as from consumption of other commodities. The utility enjoyed from driving depends on characteristics of the automobile, as well as VMT. The decision whether to own a car is integrated with the decision about VMT. Each household has exogenous income; most households also are endowed with cars. If a household has a car endowment, it chooses whether to hold or relinquish (sell or scrap) that car; if it relinquishes the car it also decides whether to purchase a different car (new or used). If a household does not have a car endowment, it chooses whether to purchase a car.

If household *i* owns car *j*, its utility can be expressed by:

(2.1) $$U_i = U_i(\boldsymbol{z}_j, M_i, x_i)$$

[10] The number of distinct cars increases over time as some unique new models become old and enter the used car fleet.

where $\mathbf{z}_j$ is a vector of characteristics of car *j*, and M_i and x_i respectively refer to household *i's* vehicle miles traveled and its consumption of the outside good. The household's utility conditional on choosing car *j* can be expressed through the following indirect utility function:

(2.2) $$V_{ij} = V_{ij}^{'} + \mu_i \varepsilon_{ij}$$

with

(2.3) $$V_{ij}^{'} = V_{ij}^{'}(y_i, r_{ij}, p_{ij}^{M}, p_{ix}, \mathbf{z}_j, \mathbf{z}_i, \mathbf{z}_{ij})$$

where

y_i	=	income to household *i*
r_{ij}	=	rental price of car *j* to household *i*
p_{ij}^{M}	=	per-mile operating cost
p_{ix}	=	price of the outside good, *x*
z_i	=	vector of characteristics of household *i*
z_{ij}	=	vector of characteristics of household *i*, interacted with characteristics of car *j*

Household income y_i is devoted toward purchasing a car (or cars[11]), car operation, and the purchase of the outside good. We treat car purchases as rentals, so that payments are spread over many years. The household budget constraint can then be written as:

(2.4) $$y_i = r_{ij} + p_{ij}^{M} M_i + p_{ix} x_i$$

If a household owns a vehicle, the stream of rental income from that vehicle is included in its income. A household that chooses to retain its existing car effectively makes a rental payment equal to its implicit rental income from that car. Income also includes the household's share of profits to new car producers, government transfers, and capital gains or

[11] In Section 4 we discuss how we allow for multiple car ownership.

losses resulting from changes in automobile prices.[12] The government transfer component of income includes revenue from the gasoline tax and adjusts as policy changes.

The operating cost p_{ij}^{M} includes the fuel cost (including gasoline taxes) as well as maintenance and insurance costs. The rental price r_{ij} accounts for depreciation, registration fees, and insurance. As indicated in expression (2.2) above, indirect utility includes the random component $\mu_i \varepsilon_{ij}$, where ε has an extreme value distribution (following the econometric model) and μ is a scale parameter. We assume the household chooses the vehicle (or vehicles) yielding the highest conditional utility, given V' and the random error. The probability that a given car j maximizes utility for household i is:

$$\text{(2.5)} \qquad \exp\left(\frac{V_{ij}^{'}}{\mu_i}\right) \Big/ \sum_j \exp\left(\frac{V_{ij}^{'}}{\mu_i}\right)$$

The indirect utility function $V_{ij}^{'}$ can be differentiated following Roy's identity to yield the optimal choice of miles traveled, M_{ij}, conditional on the purchase of each vehicle. We allow the car purchase decision of each household (which actually represents a group of similar households) to be fractional following the probabilities above. Total (or average) VMT for each household is then the demand weighted sum of the conditional VMT's determined from Roy's identity.

c. Supply of New Cars

Each of the seven producers in the model sets prices for its fleet of automobiles to maximize profits, given the prices set by its competitors and subject to fleet fuel economy

---

[12] If a household is endowed one vehicle of type j entering the period, its gain is computed as:

$$(r_j^{'} - r_j) \cdot (1 - \theta_j) + \tfrac{1}{2}(r_j^{'} - r_j)(\theta_j - \theta_j^{'})$$

where r_j and r_j' respectively denote the rental price of car j in the reference and policy-change cases, and θ_j and θ_j' represent the probability of the car's being scrapped in the two cases. The first term represents the gain in value of cars supplied in the baseline, while the second is a triangle approximating the increase or decrease in surplus associated with changes in the number of vehicles scrapped.

constraints. Thus we assume Bertrand competition. Producers face less than perfectly elastic demands for their cars: that is, two new cars of the same class can sell at different prices if produced by different producers.

The producer problem is influenced by the presence of CAFE standards. These standards require that each manufacturer's fleet-wide average fuel-economy be above a certain level in each of two general categories of cars: "light trucks" and "passenger cars." The classes in the passenger car category are non-luxury compact, non-luxury midsize, non-luxury fullsize, luxury compact, and luxury midsize/fullsize. Those in the light truck category are small truck, large truck, small SUV, large SUV/van, and minivan.[13]

In the following, the subscript k refers to the cars made by a particular manufacturer. The boldface vector $\mathbf{p}$ includes prices of the cars made by all seven manufacturers.[14] T and C denote the sets of cars (for a given manufacturer) in the light truck and passenger car categories, respectively. $\overline{e}_T$ and $\overline{e}_C$ refer to the efficiency requirements for light trucks and passenger cars and e_k is the fuel economy of car k. The profit-maximization problem solved by a given producer in our model is:

$$(2.6) \qquad \max_{\{p_k\}} \sum_k (p_k - c_k) \cdot q_k(\mathbf{p})$$

subject to:

$$\frac{\sum_{k \in C} q_k}{\sum_{k \in C} \frac{q_k}{e_k}} \geq \overline{e}_C \quad \text{and} \quad \frac{\sum_{k \in T} q_k}{\sum_{k \in T} \frac{q_k}{e_k}} \geq \overline{e}_T$$

where p_k and c_k refer to the purchase price and marginal cost, respectively, of a particular car and q_k is the demand as a function of all prices.[15] Marginal cost is assumed to be exogenous

[13] We remove a small (fixed) fraction of the largest vehicles from CAFE in order to incorporate the fact that the very largest trucks and SUV's are exempt from CAFE standards.

[14] The purchase price is the same as the present value of rental prices over the life of the car.

[15] Our treatment ignores some complexities of the CAFE regulations. The actual regulations allow for intertemporal banking and borrowing: the standard can be exceeded in one year if the firm overcomplies in

and constant. We employ data automobile markups, prices, and quantities sold, along with our estimated household demand elasticities for different automobiles, to identify the cost parameters. Details are provided in the appendix.

We must solve the constrained optimization problem for all of the firms simultaneously since the residual demand curve faced by a given firm depends on the prices set by the others. The solution method is discussed in subsection e below.

d. Used Car and Scrap Markets

i. The Used Car Market

In the model, "used car" refers to all cars that are neither new nor scrapped. For any given make and class of vehicle, ℓ, the quantity of used cars evolves according to:

$$q^U_{\ell,t+1} = (1-\theta_\ell)q^U_{\ell,t} + q^N_{\ell,t} \tag{2.7}$$

where $q^U_{\ell,t}$ and $q^N_{\ell,t}$ refer to the quantity of used and new cars of make and class ℓ available in year *t*, and θ_ℓ represents the average probability that used cars of type ℓ are scrapped.

In general, each car type, or age-make-model combination, will have a different price. The model determines the set of prices for all car types that allows market-clearing in the used-car market, that is, that causes every car to be sold. Since the demand for a given car will depend on the prices of other used cars (and on new car prices), all used car prices need to be solved simultaneously.

ii. The Scrap Market

We assume that households will scrap a car when the car's scrap value exceeds its resale value. However, each car (class-age-manufacturer combination) in our model actually

another. In addition, manufacturer can elect to pay a fine rather than meet the standards. Work in progress (Jacobsen, [2006]) addresses these issues.

represents a group of cars of varying quality and value, some of which may fall under the cutoff for scrapping even if the average car in the group does not. To allow for scrapping of some cars of a given type, we assign a scrap probability to each car. The scrap decision depends on p_j, the purchase price or resale value of a used car. This is computed as the discounted sum of future rental prices, adjusted for the possibility that a car will be scrapped before reaching each progressively older age. The household is assumed to have myopic expectations regarding future rental values; that is, it assumes that future rental values will be the same as the current-period rental values.

Once we have arrived at resale values for each used car, the scrap probability θ_j is modeled simply as:

(2.8) $$\theta_j = b_j \cdot (p_j)^{\eta_j}$$

where b_j is a scale parameter used for calibration and η_j is the elasticity controlling the change in scrap probability as the price of the car changes. Scrap rates increase with car age.

e. Solution Method

Solution of the model requires obtaining the full vector of new and used car prices that meets the following two conditions: (1) every available (not scrapped) used car has a buyer (or retainer) and (2) for every new car producer, the first-order conditions for constrained profit-maximization are satisfied. Note that the second requirement is a function of all prices, not just new-car prices. Since each household demands a range of vehicles with varying probabilities, we determine demands for a given car by aggregating across households the probability-weighted demands for that car.

The solution method embeds the used-car problem within the broader problem of solving for both used- and new-car prices. Specifically, we solve for the used-car prices that satisfy requirement (1), conditional on a set of posited prices for the new cars. We then adjust the new car prices in an attempt to meet condition (2), and solve again for used-car prices that meet requirement (1) conditional on the adjusted new-car prices. We repeat this procedure until conditions (1) and (2) are met within a desired level of accuracy. The government's revenue from gasoline taxes is returned to households according to the various "recycling" methods described in Section 5. Government revenues and household demands are mutually dependent. Hence solving the model also requires that we arrive at the level of aggregate government transfers consistent with government's receipt of tax revenue. The overall solution is a set of prices for each car that simultaneously clears all markets, and an aggregate transfer level that equals the government's revenues from the gasoline tax. To solve the multidimensional system we use Broyden's method, a derivative-based quasi-Newton search algorithm.

3. Data

Our dataset has two main components: (1) a random sample of U.S. households' automobile ownership choices from the 2001 National Household Travel Survey (NHTS) and (2) new and used automobile price and non-price characteristics from *Wards Automotive Yearbook*, The National Automobile Dealer's Association (NADA) *Used Car Guide*, and the Department of Energy (DOE) fueleconomy.org website. By merging these two types of information, we obtain an unusually rich data set, one that allows us to consider household choices among a wide range of new and used cars and that permits us to distinguish households along many important dimensions. In the appendix, we offer details on how we merged the different data sets and constructed needed variables.

a. The NHTS Sample

The 2001 NHTS consists of 26,038 households living in urban and rural areas of the United States. With the help of Department of Transportation staff, we obtained the confidential version of the NHTS data files that contained the relevant data for our analysis. For each household we have information on income, automobile holdings (by make, model, and year), and vehicle miles traveled. In addition, we have data on the household's demographic characteristics (including household size, composition, gender, education and employment status) and geographical identifiers (including the state, metropolitan statistical area, and zip code of the residence).

After cleaning the data our final sample consists of 20,429 households from the original 26,038. Table 3-1 presents major demographic statistics of our final sample.

b. The Automobile Sample

The 1983- 2002 *Wards Automobile Yearbook* provided most of the car and truck characteristics used in our analysis. Automobile characteristics include horsepower, weight, length, height, width, wheelbase, and city and highway miles per gallon (MPG) by make, model, and year for all cars and trucks sold during this period. We obtained information on car and truck prices from the National Automobile Dealer's Association (*NADA*), which publishes this information in the monthly *NADA Used Car Guide*. We used price information from the April 2001 and 2002 editions of the *Car Guide*, which we obtained in electronic format. Each edition contained the manufacturer's suggested retail price and current resale price (a weighted average of recent transaction prices) for all new and used cars and trucks dating back to 1983. As indicated in the appendix, we calculated depreciation based on changes in prices for a given car over the 2001-2002 period.

Combining information from the Wards and NADA data sets yielded a vector of prices and various automobile characteristics for roughly 4,500 automobiles distinguished by manufacturer, model, and year. We aggregated these data into the seven manufacturer categories, ten class categories, and five age categories indicated in Table 2-1. We used a weighted geometric mean formula to aggregate price and non-price characteristics within each make, class, and age category, where the weights were proportional to the holdings frequencies in the *NHTS*.

Table 3-2 displays statistics on miles per gallon and horsepower from our data. The data show significant MPG differences across classes and age categories. A new compact, for example, is 1.48 times more efficient than a large SUV. The newest compacts yield 1.47 more miles per gallon than those in the oldest age category. In contrast, the newest midsize and large SUVs are less fuel-efficient than the older models. As for horsepower, we note that most of the increases apply to compacts and full size cars. Average horsepower of compacts increased 60 percent, and average horsepower of full size cars rose 75 percent.

c. Calculation of Rental Prices and Per-Mile Operating Costs

Two important variables we must construct from our data are the automobile rental prices and per-mile operating costs (the "price per mile" variable in Section 2) for all 284 autos. The underlying inputs into these prices and costs differ by region as well as automobile type. For household *i* owning car *j*, the rental price is given by:

$$r_{ij} = D_j + 0.85 I_{ij}^A + F_{ij} + R \cdot p_j$$

where

D_j	=	depreciation in the real value of car *j*
I_{ij}^A	=	household *i*'s annual insurance costs for car *j*
F_{ij}	=	household *i*'s automotive registration fees for car *j*
R	=	real interest rate

Thus, the one-year rental price of a car is the sum of depreciation, insurance, and registration costs, plus the forgone real return on the principal value of the car.[16] For the real interest rate, *R*, we use a value of 3.89 percent, the 2001 average daily real rate on 30-year T-bills. We include insurance costs in both the rental price (associated with the choice of car) and the per-mile operating cost (associated with VMT). Representatives from State Farm Insurance suggested to us that roughly 85 percent of auto insurance premiums are fixed and independent of VMT. Hence, 85 percent of insurance costs appear in the rental price formula, with the remaining 15 percent applying to operating costs.

The per-mile operating cost, p_{ij}^M , is expressed by:

$$p_{ij}^M = \frac{p_i^{gas}}{MPG_j^*} + N_j + 0.15 I_{ij}^M$$

where

[16] If the household has purchased the car using a loan, this term can be equivalently interpreted as the interest payment on that loan.

p_i^{gas} = household i's per gallon price of gasoline
MPG_j = miles per gallon for car j [17]
N_j = per-mile maintenance and repair costs for car j
I_{ij}^M = household i's per-mile insurance costs for car j

4. Estimation of Household Ownership and Utilization Decisions

a. The Econometric Model

(i) Challenges

Two overarching concerns significantly influenced our approach to estimating household automobile demand. The first was our desire to consistently integrate car ownership and utilization decisions. Given our interest in assessing the welfare implications of gasoline taxes, linking these decisions within a utility-theoretic framework was crucial for generating consistent welfare estimates. The second concern arose from an important feature of the data – households frequently own more than one car. In the 2001 *NHTS*, 41.5 percent of households own zero or one car, another 43.6 percent own two cars, and the remaining 14.9 percent own three or more autos. This implies that many households have a potentially enormous number of auto bundles from which to choose. If, for example, there are J different cars and trucks and we consider only bundles consisting of two or fewer than two cars, there are $1+J+J(J+1)/2$ bundles that households can potentially choose. With our automobile data set consisting of 284 composite cars and trucks, there are 40,755 distinct bundles that households might choose (and this large number ignores all bundles with three or more autos).

Past efforts to integrate automobile ownership and utilization decisions have relied on reduced-form, sequentially-estimated models (Berkovec, 1985; Mannering and Winston, 1985; Train, 1986; Goldberg, 1998; West, 2004; Bento *et al.*, 2005). We adopt a consistent, utility-theoretic approach that links ownership and utilization decisions through Roy's

[17] For *MPG* we use a weighted harmonic mean of the city and highway EPA mileage rating.

identity and estimates both choice margins simultaneously. To account for different households owning different quantities of cars, we adopt a variation of Hendel (1999) and Dube's (2004) repeated discrete-continuous framework. In the context of automobile choice, the framework assumes that a household's ownership and utilization choices arise from separable choice occasions. On each choice occasion, the household makes a discrete choice of whether to own one of *J* automobiles. If an auto is chosen, the household conditionally decides how much to drive it during the year. The discrete and continuous decisions are linked through Roy's identity. To account for ownership of multiple automobiles, households have multiple choice occasions on which different automobile services may be demanded. Intuitively, different choice occasions in our framework correspond to different primary tasks or purposes for which households might demand automobile services (e.g., commuting to work, family travel, shopping excursions, or any combination thereof). We assume their number depends on the number of adults in a given household.[18]

Our approach to modeling automobile demand has advantages and disadvantages. Its main advantages are that it consistently links ownership and utilization decisions and reduces the dimension of the households' choice set on a given choice occasion to J+1 alternatives (J autos and the no auto alternative). The latter feature implies that our approach is econometrically tractable with our 284 composite auto data set. It also has the virtue of allowing for households to own any quantity of autos as long as they are given a sufficiently large number of choice occasions. Our framework's main drawback, however, is that it does not allow for interaction effects among the fleet of autos held by households. In particular, it would fail to account for car-interactions: for example, a four-person household's utility from holding a second minivan being less than holding a single minivan. To account for such interactions, one would need to treat bundles of automobiles, rather than individual cars, as the objects of choice. However, as suggested above, employing such an approach would require considerably more aggregation of cars beyond what we have pursued.[19] This would

---

[18] There is some evidence in the non-market valuation literature that the specification of the number of choice occasions, as long as it is larger than the chosen number of goods, does not have significant effects on estimated welfare measures (von Haefen, Massey, and Adamowicz, 2005). Moreover, we do not expect that it has much if any effect on the relative efficiency rankings of policies.

[19] Feng, Fullerton, and Gan's (2005) bundling approach aggregates all automobiles into one of two composites – cars and trucks.

imply less product differentiation and thus could limit our ability to account for the imperfectly competitive nature of the automobile industry. In addition, it could necessitate our putting a limit of two on the number of cars owned by a household, which would eliminate from our sample those individuals likely to be most adversely affected by gasoline taxes.

(ii) Specifics

Our repeated discrete-continuous model of automobile demand works as follows. Household i ($i = 1,\ldots,N$) is assumed to have a fixed number of choice occasions, T_i. We let T_i equal the number of adults in each household plus one.[20] On choice occasion t, household i is assumed to have preferences for car j ($j=1,\ldots,J$) that can be represented by the following conditional indirect utility function:

$$
\begin{aligned}
V_{itj} &= V_{ij}^{'} + \mu_i \varepsilon_{ijt} \\
V_{ij}^{'} &= \frac{-1}{\lambda_i} \exp\left(-\lambda_i \left(\frac{y_i / T_i - r_{ij}}{p_{ix}} \right) \right) - \frac{1}{\beta_{ij}} \exp\left(\alpha_{ij} + \beta_{ij} \frac{p_{ij}^M}{p_{ix}} \right) + \tau_{ij} \\
\alpha_{ij} &= \tilde{\alpha}_i^{\top} \mathbf{z}_{ij}^{\alpha} \\
\beta_{ij} &= -\exp(\tilde{\beta}_i^{\top} \mathbf{z}_{ij}^{\beta}) \\
\lambda_i &= \exp(\tilde{\lambda}_i^{\top} \mathbf{z}_i^{\lambda}) \\
\tau_{ij} &= \tilde{\tau}_i^{\top} \mathbf{z}_{ij}^{\tau} \\
\mu_i &= \exp(\mu_i^{*})
\end{aligned}
\tag{4.1}
$$

where $(y_i, r_{ij}, p_{ij}^M, p_{ix})$ are household i's income, rental price for the jth auto, utilization (or *VMT*) price for the jth car, and the Hicksian composite commodity price, respectively, $(\mathbf{z}_{ij}^{\alpha}, \mathbf{z}_{ij}^{\beta}, \mathbf{z}_{ij}^{\tau})$ are alternative automobile characteristics (including make, age, and class dummies) interacted with household demographics, $\mathbf{z}_i^{\lambda}$ contains just household characteristics, $(\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\lambda}_i, \tilde{\tau}_i, \mu_i^{*})$ are parameters that vary randomly across households, and ε_{ijt} contains additional unobserved heterogeneity that varies randomly across households,

[20] The 2001 *NHTS* suggests that a significant percentage of households hold an additional automobile relative to the number of adults. For the 1.84 percent of household with more autos than the one plus the number of adults, we set the number of choice occasions equal to the number of held autos.

automobiles, and choice occasions. If the household instead decides not to rent a car (i.e., automobile 0), its conditional indirect utility function is:

$$(4.2)\qquad V_{it0} = \frac{-1}{\lambda_i}\exp\left(-\lambda_i\left(\frac{y_i / T_i}{p_{ix}}\right)\right) + \phi_i^{\mathsf{T}} \mathbf{z}_i^{\phi} + \mu_i \varepsilon_{i0t}$$

where $\mathbf{z}_i^{\phi}$ and ϕ_i are individual characteristics and parameters, respectively. The rational household is assumed to choose the alternative that maximizes its utility on each choice occasion. Assuming each ε_{ijt} ($j = 0,...,J$) can be treated as independent draws from the normalized type I extreme value distribution, the probability that individual *i* chooses alternative *j* on choice occasion *t* condition on the model's structural parameters is:

$$(4.3)\qquad \Pr_{it}(j) = \frac{\exp(V_{ij}^{'} / \mu_i)}{\sum_{k=0}^{J}\exp(V_{ik}^{'} / \mu_i)}.$$

Assuming the household chooses automobile *j*, Roy's identity implies that the household's conditional *VMT* demand is:

$$(4.4)\qquad M_{itj} = \exp\left(\alpha_{ij} + \beta_{ij}\left(\frac{p_{ij}^{M}}{p_{ix}}\right) + \lambda_i\left(\frac{y_i / T_i - r_{ij}}{p_{ix}}\right)\right).$$

We assume the analyst imperfectly observes M_{itj} due to measurement error in our data.[21] The analyst observes $\widetilde{M}_{itj} = M_{itj} + \eta_{itj}$, where η_{itj} is an independent draw from the normal distribution with mean zero and standard deviation $\sigma_i = \exp(\sigma_i^{*})$. The likelihood of observing $\widetilde{M}_{itj}$ conditional on the model parameters is:

$$(4.5)\qquad l(\widetilde{M}_{itj} \mid j \text{ chosen}, j \neq 0) = \frac{1}{(2\pi)^{1/2}\sigma_i}\exp\left(-\frac{1}{2}\left(\frac{\widetilde{M}_{itj} - M_{itj}}{\sigma_i}\right)^2\right).$$

Given our assumed structure, the full likelihood of household *i*'s automobile demand conditional on $\delta = (\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\lambda}_i, \tilde{\tau}_i, \phi_i, \mu_i^{*}, \sigma_i^{*})$ is then:

$$(4.6)\qquad L_i = \prod_{t=1}^{T_i}\left[\prod_{j=0}^{J}\Pr_{it}(j)^{1_{itj}}\prod_{j=1}^{J} l(\widetilde{M}_{itj} \mid j \text{ chosen})^{1_{itj}}\right]$$

[21] Due to recall biases with the 2001 *NHTS*, it is highly probable that measurement error is present in the VMT data. We believe it is therefore appropriate to account for this source of error explicitly in our econometric model.

where 1_{itj} is an indicator function equal to one if car *j* is chosen on individual *i*'s *t*th choice occasion and zero otherwise.

b. Estimation Strategy

Past econometric efforts to model vehicle ownership and derived VMT demand decisions have used variations of Dubin and McFadden's (1984) sequential estimation strategy that accounts for the induced selectivity bias in derived VMT demand with a Heckman-like (1979) correction factor. We employ a full-information estimation approach that allows for correlations in the unobserved determinants of choice across discrete and continuous dimensions through random parameters (McFadden and Train, 2000). We assume that all parameters, $\delta = (\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\lambda}_i, \tilde{\tau}_i, \phi_i, \mu_i^*, \sigma_i^*)$, are distributed multivariate normal with mean $\bar{\delta}$ and variance-covariance matrix Σ_δ. For each observation, the same parameters enter the automobile ownership and VMT demand equations, and thereby induce correlations between the two margins of choice. Given the nonlinear nature of our likelihood function, the large number of households and sites in our data set, and the potentially large number of parameters on which we wish to draw inference, classical estimation procedures such as maximum simulated likelihood (Gourieroux and Monfort, 1996) would be exceptionally difficult, if not impossible, to implement. In light of these computational constraints, we adopt a Bayesian statistical perspective and employ a variation of Allenby and Lenk's (1994) Gibbs sampler estimation procedure that is less burdensome to implement in our application.[22]

The Bayesian framework assumes that the analyst has initial beliefs about the unknown parameters $(\bar{\delta}, \Sigma_\delta)$ that can be summarized by a prior probability distribution, $f(\bar{\delta}, \Sigma_\delta)$. When the analyst observes a set of choices **x**, she combines this choice

[22] Although the Bayesian paradigm implies a very different interpretation for the estimated parameters relative to classical approaches, the Bernstein-von Mises theorem suggests that the posterior mean of Bayesian parameter estimates, interpreted within the classical framework, are asymptotically equivalent to their classical maximum likelihood counterparts assuming a correctly specified data generating process. Following Train (2003), we interpret this result as suggesting that both approaches should generate qualitatively similar inference, and thus the analyst's choice of which to use in practice can be driven by computational convenience.

information with the assumed data generating process to form the likelihood of $\mathbf{x}$ conditional on alternative values of $(\bar{\delta}, \Sigma_\delta)$, $L(\mathbf{x} \mid \bar{\delta}, \Sigma_\delta)$. The analyst then updates her prior beliefs about the distribution of $(\bar{\delta}, \Sigma_\delta)$ to form a posterior distribution for $(\bar{\delta}, \Sigma_\delta)$ conditional on the data, $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x})$. By Bayes' rule, $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x})$ is proportional to the product of the prior distribution and likelihood, i.e., $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x}) = f(\bar{\delta}, \Sigma_\delta) L(\mathbf{x} \mid \bar{\delta}, \Sigma_\delta)/C$ where C is a constant. In general, $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x})$ will not have an analytical solution, and thus deriving inference about the moments and other relevant properties of $(\bar{\delta}, \Sigma_\delta)$ conditional on the data is difficult. However, Bayesian econometricians have developed a number of Markov Chain Monte Carlo (MCMC) procedures to simulate random samples from $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x})$ and in the process draw inference about the posterior distribution of $(\bar{\delta}, \Sigma_\delta)$.

Following Allenby and Lenk (1994), we specify diffuse priors for $(\bar{\delta}, \Sigma_\delta)$ and use a Gibbs sampler with an adaptive Metropolis-Hastings component to simulate from $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x})$. By decomposing the parameter space into disjoint sets and iteratively simulating from each set conditionally on the others, the Gibbs sampler generates simulations from the unconditional posterior distribution after a sufficiently long burn-in. The implementation details of the algorithm are described in the appendix.

One further dimension of our estimation approach is worth noting. Because of the large number of households in our data set (N = 20,429) and our desire to account for differences in automobile demand across different household types (e.g., single males, two-adult households with and without children, retired couples), we stratified the sample into 12 groups based on demographic characteristics and estimated separate models within each strata. In addition to decomposing a computationally burdensome estimation problem on a large data set into a series of more manageable estimation problems on smaller data sets, stratification allows us to better account for observable and unobservable differences among households.

c. Empirical Results

For all 12 strata, we obtain precisely estimated posterior mean values for $(\bar{\delta}, \Sigma_\delta)$.[23] Many of the parameters that are common across the 12 strata vary in magnitude considerably, suggesting that there is significant preference heterogeneity across the different subpopulations. We also find that the diagonal elements of Σ_δ are generally large, which suggests that there is considerable preference heterogeneity within each stratum as well. The latter preference heterogeneity and the highly nonlinear structure of our preference function mean that the estimated parameters do not have a simple economic interpretation. It seems more informative to examine the VMT and car-ownership elasticities implied by these parameters. We display these elasticities in Table 4.1, broken down by household and automobile types.

The first column of Table 4.1 reports short-run VMT elasticities with respect to operating costs. Across all households and cars, our model suggests a mean estimate of -0.69. Because gasoline makes up slightly less than half of per mile operating costs, this implies an average VMT elasticity with respect to gasoline price of -0.32. Graham and Glaister's (2002) concluded that past short-run estimates of VMT elasticity with respect to gasoline price average -0.15 but are quite sensitive to the data and modeling assumptions employed. Across household and different car types, we found elasticities that were higher for families with children and owners of older, luxury cars. The second column of the table reports VMT elasticities with respect to income. On average, we find estimates of around 0.62. Graham and Glaister do not report comparable estimates for VMT, but instead report short-run *gasoline* demand elasticities with respect to income of 0.35 to 0.55. In the short run, one would expect these estimates to be comparable. This elasticity was highest for families with children and owners of new luxury cars and SUVs.

The final column of Table 4.1 reports car ownership elasticities with respect to own rental price. For new cars, rental price elasticities should track purchase price elasticities if rental and purchase prices vary proportionally. Our results imply mean elasticities of -0.88

[23] Parameter estimates for each of the 12 strata are reported in the appendix.

and -1.97 for all and new cars and trucks, respectively. Luxury cars, large SUVs, and large trucks, which have the highest rental prices, have the highest rental price elasticities among automobile classes. These elasticity estimates are generally lower than recent new car purchase elasticity estimates from the industrial organization (IO) literature that range in magnitude from -3 to -4.5 (e.g., Berry, Levinsohn and Pakes, 1995; Petrin, 2003). The limited cross-sectional variation in our rental price data may play a role in explaining these differences. Berry, Levinsohn, and Pakes (1995), for example, have 20 years of new car data at the model level (roughly 200 models per year), while our data set consists of a single cross section of new and used cars aggregated to the make/class level (59 new cars, 284 total cars). The additional temporal variation in price greatly assists their efforts to reliably identify price effects.[24] By contrast, Berry, Levinsohn, and Pakes (2004), using a single cross section with 218 new cars, find lower new car elasticities that are closer in magnitude to ours.[25]

5. Simulation Results

A. Assumptions Underlying Simulation Dynamics

The simulation model generates a time path of economic outcomes over 10 years at one-year intervals. As mentioned, the model solves in each period for the market-clearing new and used car prices. The time-profile of equilibria reflects assumptions about income growth and automobile technology over the 10-year simulation interval. We assume that household incomes grow at an annual rate of one percent. In addition, we assume that automobile horsepower and weight increase at an annual rate of five percent, which is roughly consistent with historical trends. In our central case we adopt the “Path 1”

[24] Similarly, Goldberg (1995) uses five years of new car data and finds elasticities that are comparable to Berry, Levinsohn, and Pakes’.

[25] We explored the sensitivity of estimates to several alternative specifications and estimation strategies. For example, we experimented with allowing the income coefficient to vary across car classes and age groups, restricting a subset of parameters to be fixed across the sample, including alternative specific constants to account for unobserved automobile characteristics, and jointly estimating our model with five years of aggregate new car data on prices and quantities. None of these alternatives generated larger or more intuitive elasticities than those from our chosen approach.

assumptions of the National Research Council (2002) regarding improvements in fuel-economy: over a 10-year period, such improvements range from 11 percent for compacts to 20 percent for large SUVs. As part of a sensitivity analysis below, we adopt the more optimistic NAS "Path 3" assumptions regarding growth in fuel-economy. In all simulations, the pre-tax price of gasoline is $1.04 and is taken as exogenous and unchanging over time.[26]

B. Baseline Simulation

The baseline simulation offers a reference situation with which we compare the outcomes from various gasoline tax policies. Table 5-1 displays the equilibrium quantities of new and used cars under the baseline simulation. Our reference case overpredicts the size of the vehicle fleet by about 8 percent, ranging from 4 percent for midsize cars to 21 percent for luxury compacts.

C. Impacts of Gasoline Tax Increases under Alternative Recycling Methods

Here we present results from simulations of permanent increases in gasoline taxes. We start by focusing on the impacts of a tax-increase of 25 cents per gallon (other tax-increases are considered below) under the following alternative ways of recycling the additional revenues from the tax increase:

-- *"flat" recycling:* revenues are returned in equal amounts to every household

-- "*income-based" recycling:* revenues are allocated to households according to each household's share of aggregate income

-- *"vmt-based" recycling:* revenues are allocated according to each household's share of aggregate vehicle miles traveled

All of these recycling cases are consistent with a situation where the government mails rebate checks to households. Under flat recycling, the value of the rebate check is the same for all households; under the other forms of recycling, the value differs across households.

[26] Pre-existing federal taxes are $0.185 and average state taxes are $0.225.

i. Aggregate Impacts

Gasoline Consumption

Table 5-2 presents the impacts of this policy on gasoline consumption. In the short run (year 1), the percentage reduction is about 5.1 percent under flat and income-based recycling, and about 4.5 percent under VMT-based recycling. Compared with other recycling methods, VMT-based recycling gives a larger share of gasoline tax revenue to car owners, who tend to have larger income elasticities of gasoline use. As a result, there is a larger offsetting income effect on gasoline use under VMT-based recycling than under other recycling methods, and the overall reduction in gasoline consumption is smaller.

The percentage change in gasoline use is approximately equal to the percentage change in miles traveled (VMT) minus the percentage improvement in fuel economy (miles per gallon). The table shows the contributions of these two components. Most of the reduction in gasoline use comes from the reduction in VMT: there is relatively little change in fleetwide fuel economy.

In the short run, the major channel for improved fuel economy is an increase in the scrapping rate for vehicles with unusually low fuel-economy. The augmented gasoline tax raises per-mile operating costs, which makes vehicles with low fuel economy relatively less desirable, causing their demand and prices to fall and their scrap rates to rise. In the first year of the policy, an additional 160,000 used large trucks and large SUVs are scrapped. Over the longer term, average fuel-economy is influenced by changes in fleet-composition attributable to increased relative sales of new cars that are more fuel-efficient. Our simulations indicate that the percent increase in fuel-economy is larger in the long run, although the fuel-economy improvement still accounts for a small share of the overall reduction in gasoline consumption.[27]

[27] As indicated at the beginning of this section, our simulation model specifies an exogenous time-profile for the fuel-economy of new cars of each type. It thus does not allow for endogenous fuel-economy improvements for a given car model. As a result it may understate the economy-wide improvements in fuel economy from a gasoline tax.

Table 5-3 summarizes the changes in fleet composition. On impact, the higher gasoline tax occasions a shift away from cars (more cars are scrapped) and, among cars that remain in operation, a shift toward used cars (which, on average, are more fuel-efficient). In the long run, the percentage reduction in new cars is smaller. This is the case because (according to our "Path 1" assumptions) new cars become increasingly efficient relative to the overall car fleet as time passes.

Several prior studies[28] suggest that the overall reduction in gasoline consumption should be larger in the long run than in the short run, since the fleet-composition (fuel-economy) channel requires considerable time to take effect. Interestingly, our simulations indicate that the long-run reduction is smaller than the short-run reduction in percentage terms. This occurs because VMT per household falls by a smaller percentage in the long run than in the short run (see Table 5-2). This in turn stems from the fact that although in the long run there is a larger percentage reduction in the number of cars owned by the average household, there is a smaller percentage reduction in miles traveled.[29]

The results in Table 5-2 imply that each percent increase in the price of gasoline leads to a reduction of between .25 and .30 percent in the *equilibrium* demand for gasoline. It is difficult to compare this result with other studies, since other studies do not consider market equilibrium for both new and used cars, and do not consider time explicitly. However, it may be noted that Austin and Dinan (2005) report that a 30 cent per gallon increase in the gasoline tax would reduce gasoline consumption (by new cars) by 10 percent (cumulatively) over a 14-year period.

Efficiency Costs

Table 5-4 displays the efficiency cost of gasoline tax increases of 10, 25, and 75 cents per gallon. This cost is the weighted sum of the negative of each household's equivalent variation, where a household's weight is proportional to its share of the total population.

---

[28] Examples are Agras and Chapman (1999), Glaister and Graham (2000), and Johansson and Schipper (1997).

[29] In the long run, the cost of gasoline represents a smaller fraction of per-mile operating cost, a reflection of both exogenous and endogenous improvements in fleet fuel economy.

Here "cost" should be interpreted as a gross measure, since it does not net out the environmental or national security benefits stemming from the policy change.

Under flat recycling, the (gross) cost per dollar raised is $.16, $.18, and $.24, for gasoline tax increases of 10, 25, and 75 cents per gallon, respectively. The costs under the alternative recycling cases are not much different from those in the flat recycling case: the nature of recycling does not much affect the gross costs. This result requires careful interpretation. Another choice in the recycling decision is whether to return revenues in lump-sum form or instead by way of cuts in the marginal rates of prior taxes such as income or sales taxes. Prior studies have shown that returning revenues through marginal rate reductions can significantly reduce policy costs, relative to lump-sum recycling.[30] Because our simulation model does not include prior taxes (except for taxes on gasoline), we can only consider recycling through lump-sum transfers, and cannot contrast other aspects of recycling.[31]

ii. Distributional Impacts

Effects across Income Groups

Figures 5-1a and 5-1b display the impacts of a 25-cent gasoline tax increase on household income groups.[32] The distribution of impacts depends crucially on the nature of recycling. Under flat recycling, lower income groups experience a welfare improvement from the policy change, while higher income groups suffer a welfare loss. Here the lower income groups receive a share of the tax revenues that is considerably larger than their share

[30] See, for example, Goulder *et al.* (1999) and Parry and Oates (2000).

[31] The absence of prior taxes can also affect policy costs. The direction of the bias from this omission depends on the extent to which the commodity receiving the tax increase (gasoline) is a complement or substitute for taxed factors of production such as labor and capital. Previous studies indicate, in particular, that if gasoline is an average substitute for leisure, the presence of prior taxes raises the costs of a gasoline tax (or of an increase in this tax). See, for example, Goulder and Williams (2003). On the other hand, if gasoline is a sufficiently weak substitute (or relatively strong complement) for these factors, then the pre-existing taxes imply lower costs from a gasoline tax. West and Williams' (2004) empirical estimates indicate that gasoline is in fact a significant leisure-complement, which suggests that our model may bias upward the cost of a gasoline tax increase. Their study calculates the cost of an incremental increase in the gasoline tax to be about 26 cents, somewhat higher than the cost in our simulations.

[32] The pattern of impacts across households is similar for the 10-cent and 75-cent gasoline tax increases.

of gasoline tax payments. While policy discussions often refer to the potential regressivity of a gasoline tax, these simulations indicate that flat recycling more than fully offsets this potential regressivity.

Under income-based recycling the pattern of impacts is U-shaped. In this case the middle-income households experience the largest welfare loss. As indicated in Table 5-5, for these households the ratio of miles driven (or gasoline taxes paid) to income is highest; hence recycling based on income benefits these households less than other households. Only the very rich experience welfare gains under income-based recycling; these households have the lowest ratio of miles traveled (or gasoline tax paid) to income.

VMT-based recycling implies a fairly flat pattern of impacts across the income distribution, although the welfare losses are greater for higher-income households. In comparison with lower-income households, rich households drive more luxury cars, which are relatively less fuel-efficient. As a result, the ratio of gasoline taxes paid to VMT is especially large for richer households, and these households benefit least from VMT-based recycling.

Table 5-6 decomposes the welfare impacts into the various contributing factors: the change in gasoline price, the transfer (rebate) of gasoline tax revenue, the net capital gain or loss associated with policy-induced changes in car prices, and changes in profit to new-car producers. We have assumed that households own shares of new-car profits in proportion to their share of benchmark aggregate income. The table makes clear that changes in the gasoline price and the transfer are by far the most important sources of the household welfare impact. It also confirms that, depending on the type of recycling involved, the transfer may or may not offset the gasoline price impact to a particular household.

Effects along Other Demographic Dimensions

Figures 5-2a and 5-2b show VMT and policy impacts by race and income. The figures reveal two main results. First, income seems to be a more important determinant of welfare impact than race: there is greater variation in welfare impacts across income groups than across racial categories. This reflects the fact that much of the welfare impact is determined by VMT, and the differences in VMT across income groups are much larger than the VMT differences across racial groups, after controlling for income (Figure 5-2a). Second, low-income African-American households enjoy the largest gains from flat recycling, while high-income African-Americans experience the smallest losses. This is in keeping with the relatively small differences in VMT between low-income and higher-income African-American households.[33]

Figures 5-3a and 5-3b display differences in welfare impacts across states.[34] The top map displays average VMT per household from the data. The bottom map exhibits the differences in average household welfare impact. The top and bottom maps are nearly identical, indicating that benchmark VMT is a strong predictor of the welfare impact. Benchmark VMT seems to be strongly correlated with population density. Several relatively densely populated states – New York, Pennsylvania, New Jersey, and Florida – experience the smallest average welfare impact, while many of the relatively sparsely populated states – Montana, Idaho, Utah, Oklahoma, Texas, Alabama, Georgia, and South Carolina – suffer the largest adverse impacts. However, population density does not perfectly correlate with benchmark VMT or the magnitude of the impact: some sparsely populated states – Wyoming and Nevada – nevertheless have low benchmark VMT and relatively small welfare impacts.

Table 5-7 shows how impacts differ depending on the employment status of the household. Retirees fare better than younger individuals, as they tend to drive less. Households with no children also do better, for the same reason.

D. Sensitivity Analysis: Faster Technological Change and More Scrapping

[33] Although not displayed, the same pattern emerges under other forms of recycling: differences in income account for more of the variation of welfare impacts than racial differences do, and the variation in impacts between high-income and low-income African-American households is relatively small compared to the variation for other households.

[34] To generate the results in these figures, we first regressed the household welfare impacts (EVs) from the simulation on household characteristics and on the predicted baseline VMT and predicted baseline VMT squared. Next we used the coefficients from the regression, the same set of household characteristics, and household baseline VMT from the data (as opposed to predicted VMT) to get a new fitted value of EV for each household. We then aggregated this information by state.

The impacts of gasoline tax increases could well be affected by the rate of technology change in automobiles over the next decade. One aspect of faster technological improvement would be speedier growth in the fuel economy of given car models in successive years.[35] To explore this possibility, we perform an additional simulation allowing for faster improvements in fuel economy. Here we adopt the "Path 3" assumptions from the National Research Council (2002) study. In contrast with the central case, in which fuel economy increases by 11 percent for compacts to 20 percent for large SUVs, under this alternative scenario the improvements are more than twice as large (see Table 5-8 for details). The NRC study attributes the faster growth of fuel economy to supply-side cost reductions (which in turn can be attributed in part to government policies such as subsidies to R&D). Our simulations also express such a scenario. However, it should be noted that changes in the baseline time-profile of fuel-economy can also reflect changes in household preferences. Our model cannot capture such demand-side changes, since we assume a stable utility function in our econometric estimation.

Table 5-8 shows the different implications of the two technology paths. In the baseline, by year 10 average household gasoline consumption is 751.6 gallons in the fast technology improvement case. This is about 3 percent lower than in the central case baseline. Fuel-economy (miles per gallon) is about 22 percent higher in the fast-improvement case. Average VMT is also higher (by 18 percent), reflecting the lower per-mile cost of driving associated with higher fuel economy.

In the case with faster baseline fuel-economy improvements, the gasoline tax increase induces a smaller long-run percentage reduction in consumption than it does in the central case. This reflects the fact that gasoline occupies a smaller share of the household budget in this alternative scenario, implying a smaller income effect from the tax increase. The average long-run welfare impact (EV) is 18 percent smaller under the fast technology growth scenario, which is also consistent with gasoline's smaller budget share. Thus, the baseline time-profile of fuel economy significantly influences the welfare consequences of a gasoline

---

[35] Growth in fleetwide fuel economy has been promoted by the increased production and sale of hybrid vehicles. In our model, hybrid vehicles are merged with conventional cars within given manufacturer-class combinations (Toyota compacts). We are considering splitting out hybrids in future work. To estimate demands for hybrids, we may need to supplement our revealed-preference data with stated-preference information, since hybrids were introduced in the automobile fleet in 2001, the year corresponding to our benchmark data. Today they represent about four percent of the compact car fleet.

tax. For any given baseline scenario, introducing the 25 cent gasoline tax has a significant impact on welfare. But differences in assumptions about the baseline have larger implications for welfare than whether the gasoline tax increase is introduced.

We also perform a simulation in which we double increase the scrap elasticity η_j to -6.0 from its central value of -3.0. With this change, the gasoline tax causes a somewhat larger reduction in gasoline use in the short run, reflecting a higher scrapping rate: with the higher scrap elasticity, the policy change causes 22 percent more cars to be scrapped compared with the policy under the central case. While the higher scrap elasticity implies a larger policy impact on gasoline consumption in the short run, it has little influence on the policy impact in later years.

6. Conclusions

This paper has examined the impacts of gasoline tax increases with a model that aims to capture simultaneously several elements of the markets that importantly influence the supply and demand for gasoline. We consider together the markets for new, used, and scrapped vehicles, as well as the imperfectly competitive nature of the automobile industry. Linking the three markets enables us to account for the penetration of the car fleet by new cars and thereby assess how the impacts of policy interventions evolve through time. We also address the considerable range of car choices in a fairly high-dimensional discrete choice model. Parameters for the household demand side of the model are estimated econometrically using a one-step procedure that integrates individual choices for car ownership and miles traveled, thereby yielding consistent welfare measures. Finally, we allow for the considerable heterogeneity among car owners, which enables us to explore the distributional impacts of policy changes along many important dimensions.

We find that each cent-per-gallon increase in the price of gasoline reduces the equilibrium gasoline consumption by about .2 percent. The reduction in demand mainly reflects reduced miles traveled by car owners; shifts in demand from low to high miles-per-gallon vehicles appear much less important. The gasoline tax's marginal excess burden (excluding external benefits) per dollar of revenue raised ranges from about \$.15 for a 10-

cent tax-increase to $.25 for a 75-cent increase. Taking account of revenue-recycling (and disregarding external benefits), the impact of a 25-cent gasoline tax increase on the average household is about $30 per year (2001 dollars).

The distributional impacts of the gasoline tax differ dramatically under the three revenue-recycling approaches we considered. Under flat recycling, the average household in each of the bottom four income deciles experiences a welfare gain from a gasoline tax increase. The gain to the average household in the lowest income decile would be equivalent to about $125. This suggests that a single-rebate-check approach to recycling would more than eliminate (for the average household within a given income group) the potential regressivity of a gasoline tax increase. On the other hand, if revenues are recycled in proportion to income, only very poor households (those in the lowest decile) and very rich households (those in the highest) stand to gain. The different impacts of the various recycling methods largely reflect differences across the income distribution in car use (VMT). However, household income does not perfectly correlate with VMT and other important determinants of the welfare impacts: controlling for income, we find significant differences in impacts across racial categories and regions of residence.

The framework presented here has considerable potential to address other automobile-related policies, including tightening of CAFE standards and subsidies to retirements of low-mileage (or high-polluting) automobiles. We plan to investigate these policies in future work, examining impacts not only on gasoline consumption but on automobile-generated pollution as well.

Some limitations in our model deserve mention. It treats the time-path of fuel-economy for individual car types as exogenous. In addition, it deals only implicitly (through exogenous improvements in fuel-economy) with the advent of cars involving new technologies, such as hybrids. Finally, our estimation of automobile demands is based on cross-section data; developing a panel would offer much additional useful information. We will be exploring ways to improve the model along each of these dimensions.

References

Agras J., and D. Chapman, (1999)“The Kyoto Protocol, CAFE Standards, and Gasoline Taxes,” *Contemporary Economic Policy* 17:296-308.

Allenby, G., and P. Lenk (1994), “Modeling Household Purchase Behavior with Logistic Normal Regression,” *Journal of the American Statistical Association*, 89:669-679.

Austin, D. and T. Dinan (2005), “Clearing the Air: The Costs and Consequences of Higher CAFE Standards and increased Gasoline Taxes,” *Journal of Environmental Economics and Management* 50:562-582.

Bento, A., M. Cropper, M. Mobarak and K. Vinha (2005), “The Effects of Urban Spatial Structure on Travel Demand in the US,” *Review of Economics and Statistics* 87:466-478.

Berkovec, J. (1985), “New Car Sales and Used Car Stocks: A Model of the Automobile Market,” *RAND Journal of Economics* 16:195-214.

Berry, S, J. Levinsohn, and A. Pakes (1995), “Automobile Prices in Market Equilibrium,” *Econometrica* 63:841-890.

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

Dube, J. (2004), “Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks,” *Marketing Science* 23:66-81.

Dubin, J., and D. McFadden (1984), “An Econometric Analysis of Residential Appliance Holdings and Consumption,” *Econometrica* 52:345-362.

Feng, Y., D. Fullerton, and L. Gan (2005), “Vehicle Choices, Miles Driven, and Pollution Policies,” NBER Working Paper No. 11553, Cambridge, MA.

Glaister, S. and D. Graham, (2002) “The Demand for Automobile Fuel: A Survey of Elasticities,” *Journal of Transport Economics and Policy* 36:1-25.

Goldberg, P.K. (1995), “Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry,” *Econometrica* 63:891-951.

Goldberg, P.K. (1998), “The Effects of the Corporate Average Fuel Efficiency Standards in the US,” *Journal of Industrial Economics* 46:1-33.

Goulder, L., I Parry, R. Williams III and D. Burtraw, (1999), "The Cost-Effectiveness of Alternative Instruments for Environmental Protection in a Second Best Setting," *Journal of Public Economics* 72:329-360.

Goulder, L and R. Williams III (2003), "The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and Practical Solutions" *Journal of Political Economy* 111:898-927.

Gourieroux, C., and A. Monfort (1996), *Simulation-Based Econometric Methods*. Oxford University Press, New York, NY.

Hausman, J. and W. Newey (1995) "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica* 63:1445-1476.

Heckman, J. (1979), "Sample Selection Bias as a Specification Error," *Econometrica* 47:153-161.

Hendel, I. (1999), "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies* 66:423-446.

Hensher, D., F. Milthorpe, and N. Smith (1990), "The demand for vehicle use in the urban household sector - theory and empirical-evidence," *Journal of Transport Economics and Policy* 24:119-137.

Jacobsen, Mark R. (2006), "Impacts of Changes in CAFE Standards in an Econometrically Based Simulation Model," unpublished manuscript, Stanford University.

Johansson O, and L. Schipper, (1997), "Measuring the Long-Run Fuel Demand for Cars," *Journal of Transport Economics and Policy* 31:277-292.

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

National Research Council (2002), "Effectiveness and Impact of Corporate Average Fuel Economy Standards," National Academy Press.

McFadden, D., and K.E. Train (2000), "Mixed MNL for Discrete Response," *Journal of Applied Econometrics* 15:447-470.

Parry, I. and K. Small (2005), "Does Britain or The United States Have the Right Gasoline Tax?" *American Economic Review* 95:1276-1289.

Parry, I, C. Fischer, Carolyn and W. Harrington (2004) "Should Corporate Average Fuel Efficiency Standards be Tightened?," *Resources for the Future Discussion Paper* 53.

Parry, I, and W. Oates (2000), “Policy Analysis in the Presence of Distorting Taxes,” *Journal of Policy Analysis and Management*, 19(4):603-13.

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

Poterba, J. (1991), “Is the Gasoline Tax Regressive?” in D. Bradford, ed., *Tax Policy and the Economy 5*:145-164.

Schmalansee and Stocker (1999), “Household Gasoline Demand in the United States”. *Econometrica* 67:645-62.

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

Train, K.E. (2003), *Discrete Choice Analysis with Simulation.* Cambridge University Press, Cambridge, UK.

von Haefen, R.H., D.M. Massey, and W.L. Adamowicz (2005), “Serial Nonparticipation in Repeated Discrete Choice Models,” *American Journal of Agricultural Economics* 87:1061-1076.

West S. and R. Williams III (2005), "The Cost of Reducing Gasoline Consumption” *American Economic Review* 95:294-299.

West, S.E. (2004), “Distributional Effects of Alternative Vehicle Pollution Control Policies,” *Journal of Public Economics* 88:735-757.

Appendix for
"Distributional and Efficiency Impacts of Increased U.S. Gasoline Taxes"

A Data

Our automobile data set has two main components: (1) a random sample of U.S. household's complete automobile choices from the 2001 *National Household Travel Survey* (*NHTS*); and (2) new and used automobile price and non-price characteristics from *Wards Automotive Yearbook*, *The National Automobile Dealer's Association* (*NADA*) *Used Car Guide*, and the U.S. Environmental Protection Agency and Department of Energy's (*EPA/DOE*) *fueleconomy.org* web site. We refined and augmented these data with additional information from a number of sources summarized in Table 1.

The next section discusses the specific data we use from each source in Table 1. Section 2 follows with a summary of how we merged the data into our final data set. We conclude in section 3 with a summary statistics from the final data set.

1 Data Sources

1.1 2001 National Household Travel Survey (NHTS)

Our household automobile choice data comes from the 2001 *National Household Travel Survey* (*NHTS*), the most recent and comprehensive survey of U.S. automobile demand. Sponsored by three agencies within the U.S. Department of Transportation (DOT) and conducted between March 2001 and May 2002, the survey collected automobile demand and socio-demographic data from a random sample of U.S. households. A random digit dialing protocol was initially used to screen and elicit household participants, and follow-up household and personal phone surveys as well as written diaries collected all relevant information. The survey response rates ranged from 41.2 percent at the household level to 32.2 percent at the personal (household members at least 18 years old) level. Although somewhat low, these response rates are not surprising given the survey's length and complexity, and the *NHTS* includes sampling weights that adjust for nonresponse bias. In total, responses from 26,038 households are included in the *NHTS*.

With help from DOT staff, we obtained the confidential version of the *NHTS* data files that contained the following information relevant to our analysis:

1) Household income
2) All automobile holdings by make, model, and year[1]
3) Vehicle Miles Traveled (VMT) for each automobile over the past 12 months
4) Household and individual level socio-demographic data (e.g., household size and composition, sex, education, employment status, etc.)
5) Geographic data (state, Metropolitan Statistical Area (MSA), and zip code)

Although the quality of the *NHTS* data was generally high, we found it to be deficient in a few important dimensions. About 14 percent of households (3,723 in total) did not supply complete income data. Among those who did, some older households reported incomes that were implausibly low in our judgment. To address these data deficiencies, we first imputed 1999 Census zip code level median household incomes adjusted to 2001 dollars using the U.S. CPI. We then adjusted the incomes of all households with one or more senior citizens (adults 65 and above) to be at least the 2001 average Social Security payment for senior citizens ($10,224 per senior citizen). In total, these latter income adjustments affected 1,006 households in our sample.

Other limitations with the data forced us to drop observations from our analysis. The *NHTS* data files do not always contain sufficient information to identify all cars or the characteristics of cars households drive. Our treatment of these cars and households depended on whether the unidentified cars' VMTs were more than 1000 miles. If the unidentified car's VMT was above 1000 miles, the household was dropped from our analysis; otherwise, the household was retained and the unidentified car ignored. 2,542 households were dropped from our analysis as a result.

[1] The survey also collected detailed information on motorcycle and motor home holdings. Because these vehicles generally serve different household purposes than automobiles, we choose to ignore them in our analysis.

A related difficulty arose with households who held autos produced on or before 1983. As we discuss below, our automobile price and characteristics data did not cover these older autos. Similar to unidentified autos, our treatment of these older cars and households depended on whether the unidentified cars' VMTs were more than 1000 miles. If the older car's VMT was above 1000 miles, the household was dropped from our analysis; otherwise, the household was retained and the older car ignored. 2,332 households were consequently dropped from our analysis.

Finally, VMT data was sometimes incomplete or implausible. The *NHTS* collected annual VMT information through a veriety of channels: 1) direct questioning (i.e., "what was the total VMT for this automobile in the past 12 months?"), 2) extrapolations based on changes in reported odometer readings over a two month period, and 3) extrapolations based on detailed one-day travel diaries. Based on household responses, the *NHTS* reports two VMT estimates for each automobile: *annmile* and *bestmile*. The former relies exclusively on the direct questioning approach to eliciting VMT and the later uses regression techniques and arbitrary judgments to combine all three. Both variables have missing values, but *bestmile*, because it is based on more information, has fewer. Values for both variables range from 1 to 200,000, but our judgment was that estimates below 100 miles or above 60,000 miles were implausible, and therefore we recoded them as missing. In our analysis we use *annmile* as our VMT measure whenever available and *bestmile* otherwise. If both are missing, we dropped the corresponding household from our analysis. This protocol resulted in us losing an additional 719 households. We also recorded as missing all VMTs for a given household if its average VMT per adult was greater than 60,000 miles. This last step resulted in 16 additional households being dropped.

As a result of these cleaning procedures, 20,429 of the original 26,038 households remained in our estimation sample. Our estimation results are therefore based on 78.8 percent of the *NHTS* sample, a relatively large percentage in our view. To make our usable subsample from the *NHTS* broadly representative of the general population, we adjusted the sampling weights based on geographic and socio-demographic criteria accordingly.

1.2 NBER Tax Calculator

Because a significant portion of household income is subject to federal, state, and local taxation, we used the *NBER*'s publicly available software package *TAXSIM, version 5.1* to calculate each household's after-tax income. *TAXSIM* requires several pieces of information as input (e.g., tax year, state of residence, marital status, number of children, wages of the primary tax filer and his/her spouse, pensions, unemployment compensation), and not all of this information was available from the *NHTS*. We therefore made the following set of assumptions when calculation after tax income:

1) The tax year was taken as the year of the household interview (either 2001 or 2002).
2) Marital status was derived from the marital status of the main *NHTS* survey respondent. Our sense was the main *NHTS* survey respondent was the head of household in most cases, but we recognize that the respondent might be any adult 18 years or older in the household.
3) In households with more than one working adult, the income of each adult is frequently missing; only total household income is reported. In these cases, we assume that the tax filer and spouse's income are 60 and 40 percent of total household income, respectively.
4) If both tax filer and his/her spouse are 65 and above, their total income is treated as pensions.
5) If all adults in the household are unemployed but less than 65, their total income is treated as unemployment compensation.

1.3 Wards Automotive Yearbook

The 1983-2002 *Wards Automobile Yearbook* provided most of the car and truck characteristics used in our analysis. Characteristics obtained from *Wards* include horsepower, weight, length, height, width, wheelbase, and city and highway miles per gallon (MPG) by make, model, and year for all cars and trucks sold during this time period. The data was scanned into electronic format and carefully checked for errors and inconsistencies. Some missing characteristics were imputed through regression analysis.

1.4 NADA Used Car Guide

The National Automobile Dealer's Association (*NADA*) publishes monthly the *NADA Used Car Guide*, a detailed summary of new and used car and truck prices. We obtained from *NADA* the April 2001 and 2002 editions in electronic format. Each edition contained the manufacturer's suggested retail price and current resale price (a weighted average of recent transaction prices) for all new and used cars and trucks dating back to 1983. After deflating 2002 resale car and truck prices to 2001 dollars with the CPI, we differenced the 2001 and 2002 resale prices ($(P_j^{2001}$ and P_j^{2002}, respectively$)$) for car *j* to construct estimates of real depreciation D_j for a particular make, model, and year, i.e.,

$$D_j = P_j^{2001} - P_j^{2002} / (1 + CPI) \ .$$

Although we generally found estimates consistent with a 20 percent real depreciation rate and that varied in intuitive ways, careful inspection revealed that some missing and implausible estimates arose. In these cases, regression analysis was used to generate imputations.

1.5 EPA/DOE's Fueleconomy.org Web Site

The U.S. Environmental Protection Agency (*EPA*) and Department of Energy (*DOE*) jointly sponsor the web site fueleconomy.org which contains city and highway miles per gallon (MPG) data by make, model, and year for automobiles sold in the U.S. between 1985 to 2002. This data was helpful in checking for inconsistencies with the *Wards'* MPG data. The site also contains information on *EPA/DOE* car classification which was helpful for aggregating similar autos in our analysis.

1.6 Maintenance and Repair Costs

After numerous discussions with academic and government transportation researchers, we could not identify a comprehensive and up-to-date data set of average maintenance and repair (M&R) costs for new and used automobiles. The best data on M&R costs we could find came from *Edmunds.com* and *AAA*. For recent makes and models, the *Edmunds.com* web site reports ownership cost estimates (called the "True Cost to Own") for the initial five years of a car's use that contain separate M&R components. The maintenance component includes both scheduled (i.e., factory recommended items) and unscheduled (tires, brakes, battery, etc.) maintenance. Repair represents average repair costs not covered by the manufacturer's warranty. Both the repair and maintenance estimates assume 15,000 miles driven per year. Similarly, *AAA* annually reports per mile M&R costs (defined similar to *Edmunds*) for four representative new automobiles.

Due to the limited scope of the *Edmunds.com* and *AAA* data, we were forced to make strong and somewhat arbitrary judgments about how M&R costs in general relate to our limited set of estimates. Both the *Edmunds'* and *AAA* data consistently suggested that per mile M&R costs were roughly proportional to per mile gasoline costs. In general, new autos had slightly lower M&R costs relative to their gas costs, while older pre-1995 autos had slightly higher costs. We therefore decided to set the per mile M&R costs to 90 percent of per mile gas costs for 2001-2002 autos, 95 percent for 1999-2000 autos, 100 percent for 1995-1998 autos, 105 percent for 1990-1994 autos, and 110 percent for pre-1990 autos.

1.7 ACCRA Regional Price Data

Our gas price and regional cost of living index (COLI) data came for the American Chamber of Commerce's 2000-2002 *ACCRA* data base. Every quarter, the American Chamber of Commerce publishes per-gallon gas prices and summary COLI indexes for over 300 Metropolitan Statistical Areas (MSAs) and rural communities. This geographic resolution permitted us to link fairly precise measures of gasoline prices and regional COLIs to each household in the *NHTS*. To account for the relatively small number of missing prices, regression techniques were again used.

1.8 NAIC & State Farm Insurance Data

We develop insurance cost estimates that vary by state, vehicle class and year based on published data from the National Association of Insurance Commissioners (*NAIC*) and unpublished data from the *State Farm* Insurance Company. *NAIC* publishes periodically state level average insurance expenditures and premiums for personal automobile insurance, and we use their 2001 published estimates. *State Farm* supplied us with unpublished adjustment factors that allowed us to scale these estimates upwards or downwards to account for differences in automotive class and age.

1.9 FHWA's Auto Registration Fees Data

The Federal Highway Administration (*FHWA*) periodically collects and publishes state level automotive registration fee data. We use their 2001 data that we accessed on June 1, 2004 from the web at http://www.fhwa.dot.gov/ohim/hwytaxes/2001/pt11.htm. Although useful in terms of identifying differences in registration fees across states, the data is limiting in the sense that it does not contain information on new and used car taxes or the costs of environmental emissions testing.

2 Merging the Alternative Data Sources into the Final Data Set

2.1 Merging the Car Characteristics & Constructing Aggregate Automobiles

As discussed above, our new and used car prices and characteristics come from three primary sources – *Wards*, *NADA*, and *EPA/DOE*. Merging these data sources together proved challenging because no common vehicle identification code was present in each data set and the levels of coverage and aggregation for each make, model, and year varied considerably across the data sets.[2] Moreover, the description of automobiles in the *NHTS* is highly aggregated – we only know the make, model, and year of a given automobile,[3] and nothing about its engine size

---

[2] An example about the nature of the problem we confronted may be instructive. Consider a 1995 Honda Civic. In one data set, this make, model, and year triplet might be reported as two separate vehicles – a 1995 Honda Civic Coupe and a 1995 Honda Civic Sedan. In another data set it might be reported as three vehicles – the 1995 Honda Civic CRX, the 1995 Honda Civic DX, and the 1995 Honda Civic S. In the third data set it may be reported as four vehicles - the 1995 Honda Civic CRX 2 Door, the 1995 Honda Civic CRX 4 Door, the 1995 Honda Civic DX manual, and the 1995 Honda Civic DX automatic. Notice that none of the car descriptions across the three data sets match perfectly. They not only differ in terms of their degrees of aggregation but also in terms of their coverage.

[3] In some cases the *NHTS* does not even report the specific model a household owns but rather a group of models that contains the specific model owned. In general the models fell within a common class (e.g., the Volkswagon Golf and Cabriolet). This data limitation required further aggregation across automobile models.

(e.g., 4 or 6 cylinders), transmission (automatic or manual), or body shape (coupe, sedan, hatchback, or wagon). To address these data limitations, we separately collapsed the price and non-price characteristics for all vehicles with a common make, model, and year into a single vector of characteristics using an unweighted geometric mean formula before merging the alternative data sets. Once this initial aggregation was performed, merging the three data sets by common make, model, and year was feasible, although regression analysis was necessary to fill in a relatively small number of missing values. One limitation with the merge was that although *Wards* and *NADA* had coverage for the same years (1983-2002), the *EPA*/*DOE*'s coverage was only from 1985 to 2002. Since only the MPG and class variables from the *EPA*/*DOE* data set were used in our final analysis and the *Wards* data also contained MPG estimates, we relied exclusively on *Wards'* MPG data and assumed the car classes for particular makes and models in 1985 were the same in 1983 and 1984.

After the merge was completed, roughly 4,500 distinct make, model, and year combinations remained in our data set. Including such a relatively large number of choice alternatives in our econometric model was not feasible and thus additional aggregation was required. We therefore stratified cars into seven make categories (Ford, Chrysler Daimler, GM, Honda, Toyota, other East Asian, and European), ten class categories (non-luxury compact, non-luxury midsize, non-luxury fullsize, luxury compact, luxury midsize/fullsize, small truck, large truck, small SUV, large SUV/van, and minivan),[4] and five age categories ('01-02, '99-'00, '95-'98, '90-'94, and '83-'89). We used a weighted geometric mean formula to aggregate price and non-price characteristics within each make, class, and age category, where the weights were proportional to the holdings frequencies in the *NHTS*.[5] This approach to aggregation resulted in a total of 284 composite cars being generated.[6]

---

[4] Compacts with manufacturer's suggested retail price (MSRP) greater than $31,000 (in $2001 $s) were treated as luxury automobiles. Similarly, midsize and fullsize autos with MSRPs greater than $35,000 were treated as luxury autos. Large trucks and SUVs were assumed to have curbside weights greater than 4,000 and 4,250 pounds, respectively.

[5] Using the simple frequency weights from the *NHTS* implied that most make, model, and year combinations did not enter into the aggregate autos. To avoid this we "smooth" the weights as follows. One-third of the weight for a particular make/model/year combination remained with the specific make/model/year, another third was allocated evenly across years where similar makes and models were produced, and the final third was distributed evenly across similar makes and classes within the particular year the car was produced. The sum of these "smoothed" weights is the weight used to construct aggregate cars.

[6] Note that 66 of the possible 350 make, class, and age strata had no cars or trucks in them and were not included in our analysis.

Table 1
Data Sources

Source	***Description***	***Main variables***
NHTS	Federal Highway Administration's (FHWA) 2001 *National Household Travel Survey,* confidential files	Demographic & geographic household characteristics, automobile characteristics (e.g., make, model, year, VMT)
NBER	*TAXSIM, version 5.1*	Federal tax, state tax, & FICA for every household in NHTS
Ward's	*Ward's 2000-2002 Automotive Yearbook*	Car characteristics (e.g., horsepower, length, weight, height, width, wheelbase) for 1979-2002 make/models
EPA/DOE	Department of Energy & Environmental Protection Agency's www.fueleconomy.org web site	EPA city & highway MPGs, car class, engine size (liters, cylinders) for 1985-2002 make/models (accessed on June 1, 2004)
NADA	2001 & 2002 *National Automobile Dealers' Association Used Car Guide*	Used car prices and Manufacture's suggested retail prices (MSRP) for all 1982-2002 make/models in April & December of 2001-2002
Edmunds.com and AAA	Edmunds "True Cost to Own" data accessed at http://www.edmunds.com, and AAA's "Your Driving Costs" annual publication, 1990-present	Maintenance & repair data (accessed at Edmunds.com on June 15, 2004)
ACCRA	American Chamber of Commerce Researchers Association's *Regional Cost of Living Index*	2001-2002 composite cost of living & gas price indexes by metropolitan statistical area (MSA)
NAIC	National Association of Insurance Commissioners' *2001 State Average Expenditures & Premiums for Personal Automobile Insurance*	2001 average state level insurance expenditures
State Farm	Personal communication with State Farm Insurance's national public relations office	Adjustment factors for vehicle class & year from State Farm
FHWA	Federal Highway Administration, Office of Highway Policy Information's web site, http://www.fhwa.dot.gov/policy/ohpi/	2001 State level auto registration fees (accessed on June 1, 2004)

B *Demand Estimation*

1 *Estimation Algorithm*

Following Allenby and Lenk (1994), we specify diffuse priors for $(\bar{\delta}, \Sigma_\delta)$ and use a Gibbs sampler with an adaptive Metropolis-Hastings component to simulate from $f(\bar{\delta}, \Sigma_\delta \mid \mathbf{x})$. By decomposing the parameter space into disjoint sets and iteratively simulating each set conditionally on the others, the Gibbs sampler generates simulations from the unconditional posterior distribution after a sufficiently long burn-in.

We assume the following diffuse priors for $\bar{\delta}$ and Σ_δ:

$$\text{(1a)} \qquad \begin{aligned} &\bar{\delta} \sim N(\delta^{FP}, \tau I_k) \\ &\Sigma_\delta \sim IW(k, I_k), \end{aligned}$$

where $N(\cdot)$ and $IW(\cdot)$ denote the multivariate normal and inverse Wishart distributions, respectively, δ^{FP} are the fixed parameter maximum likelihood estimates, τ is a scalar chosen such that $1/\tau$ approaches zero, k is the dimension of δ, and I_k is a k-dimensional identity matrix. These priors, in combination with our assumed data generating process, imply the following conditional posterior distributions for $\bar{\delta}$ and Σ_δ as well as the individual specific δ_i:

$$\text{(2a)} \qquad \begin{aligned} f(\bar{\delta} \mid \delta_1, \ldots, \delta_N, \Sigma_\delta, \mathbf{x}_1, \ldots, \mathbf{x}_N) &\propto N(\bar{B}, \Sigma_\delta / N) \\ f(\Sigma_\delta \mid \delta_1, \ldots, \delta_N, \bar{\delta}, \mathbf{x}_1, \ldots, \mathbf{x}_N) &\propto IW\left[k + N, (kI_k + N\bar{S})/(k+N)\right] \\ f(\delta_i \mid \bar{\delta}, \Sigma_\delta, \mathbf{x}_i) &\propto L_i(\mathbf{x}_i \mid \delta_i) \times n(\delta_i \mid \bar{\delta}, \Sigma_\delta) \quad \forall i, \end{aligned}$$

where $L_i(\mathbf{x}_i \mid \delta_i)$ is the conditional likelihood function from equation (6) for individual i, $n(\cdot)$ is the normal density function, and

$$\begin{aligned} \bar{B} &= N^{-1} \sum_i \delta_i \\ \bar{S} &= N^{-1} \sum_i (\delta_i - \bar{\delta})^{\mathsf{T}} (\delta_i - \bar{\delta}). \end{aligned}$$

The Gibbs sampling algorithm proceeds by iteratively drawing from the conditional distributions in (2a), with each draw being made conditional on the remaining parameters' most recent draws. As Train (2003) describes, simulating from the multivariate normal and inverse Wishert distributions is relatively straightforward. However, simulating from δ_i's posterior distribution is more complex and requires an adaptive Metropolis-Hastings algorithm (Chib and Greenberg, 1995). Thus iteration *s* of the Gibbs sampler involves the following steps:

1) Simulate $\bar{\delta}^s$ from $N(\bar{B}^{s-1}, \Sigma_\delta^{s-1}/N)$. To initialize the algorithm, set $\Sigma_\delta^0 = kI_k$ and $\delta_i^0 = \bar{B}^0 = \delta^{FP}$, $\forall i$.
2) Simulate Σ_δ^s from $IW\left[k+N, (kI_k + N\bar{S}^s)/(k+N)\right]$ where $\bar{S}^s = T^{-1}\sum_i (\delta_i^{s-1} - \bar{\delta}^s)^\mathsf{T}(\delta_i^{s-1} - \bar{\delta}^s)$.
3) Simulate δ_i^s for each observation using one iteration from the following Metropolis-Hastings algorithm:
    a. For each observation, simulate a candidate vector $\tilde{\delta}_i^s$ from $N(\delta_i^{s-1}, r^{s-1}\Sigma_\delta^s)$, where r^{s-1} is a constant. To initialize the sequence, set r^0=0.1.
    b. For each observation, construct the following statistic:
    $$\chi_i^s = \frac{l(\mathbf{x}_i \mid \tilde{\delta}_i^s)n(\tilde{\delta}_i^s \mid \bar{\delta}^s, \Sigma_\delta^s)}{l(\mathbf{x}_i \mid \delta_i^{s-1})n(\delta_i^{s-1} \mid \bar{\delta}^s, \Sigma_\delta^s)}.$$
    If $\chi_i^s \geq U_i^s$ where U_i^s is a uniform random draw, accept the candidate random parameters, i.e., $\delta_i^s = \tilde{\delta}_i^s$. Otherwise, set $\delta_i^s = \delta_i^{s-1}$.
    c. Gelman et al. (1995) argue that the Metropolis-Hastings algorithm for the normal distribution is most efficient if the acceptance rate of candidate parameters averages between 0.23 and 0.44. Therefore, we set r^s=(1.01)r^{s-1} if the sample's proportion of accepted candidate parameter values is less than 0.3. Otherwise, set r^s=(0.99)r^{s-1}.
4) Iterate.

After a sufficiently long burn-in, this algorithm generates random draws from the posterior distributions of δ_i, $\overline{\delta}$, and Σ_δ. In practice, the burn-in length necessary to achieve convergence (i.e., random draws from the posterior distributions) is difficult to ascertain. However, our experience has been that the Gibbs sampler algorithm is relatively fast even in our large choice set application, and thus the analyst can cheaply add burn-in iterations if convergence is in doubt. Finally, because the Gibbs sampler induces serial correlation in δ_i, $\overline{\delta}$, and Σ_δ, we only use each 10^{th} simulation after the burn-in to construct distributional summary statistics and welfare measures.

Due to the large number of households in our data set (N = 20,429) and our desire to account for differences in automobile demand across different household types, we stratified the sample into 12 different groups based on demographic characteristics and estimated separate models within each strata. The stratification criteria and resulting strata sizes are summarized in Table 1.

2 Empirical Results

All posterior mean and corresponding variance parameter estimates for the 12 different strata are summarized in Tables 2 through 5. These estimates were generated with a total of 40,000 iterations of our Gibbs sampler estimation algorithm where we treated the first 30,000 iterations as burn-in and used every 10^{th} iteration thereafter to construct the reported estimates.

Table 1a
Strata Definitions

Strata	*Initial Size*	*Description*
1	1167	1 male adult, no children, not retired
2	1609	1 female adult, no children, not retired
3	2096	1 adult, no children, retired
4	1450	2+ adults w/ average age ≤ 35, no children, not retired
5	1722	2+ adults w/ average age > 35 & ≤ 50, no children, not retired
6	1846	2+ adults w/ average age > 50, no children, not retired
7	1897	2+ adults w/ average age ≤ 67, no children, retired
8	1730	2+ adults w/ average age > 67, no children, retired
9	1777	1+ adults w/ youngest child < 3 years old
10	1562	1+ adults w/ youngest child 3-6 years old
11	1765	1+ adults w/ youngest child 7-11 years old
12	1808	1+ adults w/ youngest child 12-17 years old
Total	20429	

Adults are at least 18 years old. Unweighted geometric mean formula used to calculate average adult age. Retirement status is self-reported. Cleaned size results from dropping households that have more cars than adults + 1 or that can not afford cars they are observed to purchase.

Table 2a
Posterior Mean Parameter Estimates – Strata 1-6

Strata	***#1***		***#2***		***#3***		***#4***		***#5***		***#6***	
	Mean	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*
α parameter												
Age category #1 – 2001-02	1.95	0.11	1.99	0.05	1.98	0.08	2.34	0.05	2.25	0.06	2.35	0.05
Age category #2 – 1999-00	2.18	0.08	2.14	0.06	2.11	0.07	2.43	0.05	2.30	0.06	2.45	0.05
Age category #3 – 1995-98	2.39	0.07	2.29	0.08	2.28	0.08	2.51	0.04	2.30	0.04	2.51	0.04
Age category #4 – 1990-94	2.16	0.06	2.04	0.04	1.97	0.06	2.20	0.05	2.13	0.05	2.16	0.04
Age category #5 – 1983-89	1.66	0.04	1.37	0.10	1.46	0.05	1.41	0.07	1.54	0.09	1.52	0.07
Horsepower (HP) / weight	-0.08	0.01	0.01	0.01	0.01	0.00	0.00	0.01	0.02	0.00	0.00	0.00
(HP / weight)*(avg adult age)	-0.36	0.65	-7.58	0.52	-6.74	0.37	-5.56	0.77	-6.60	0.56	-4.62	0.46
# of females	-	-	-	-	-0.46	0.05	-0.09	0.11	-0.02	0.06	-0.31	0.09
# of workers	-	-	-	-	-	-	0.08	0.07	0.21	0.06	0.39	0.04
β parameter												
Compact	1.56	0.04	1.43	0.05	1.54	0.03	1.57	0.07	1.77	0.03	1.81	0.04
Luxury compact	2.74	0.16	2.43	0.09	3.37	0.10	2.32	0.13	2.93	0.07	2.07	0.04
Midsize	1.67	0.06	1.58	0.03	1.44	0.07	1.77	0.05	1.77	0.04	1.64	0.04
Fullsize	2.25	0.15	2.01	0.05	1.36	0.04	2.25	0.16	1.89	0.05	1.79	0.06
Luxury midsize/fullsize	2.79	0.11	2.71	0.13	1.53	0.05	2.81	0.17	2.12	0.06	1.74	0.06
Small SUV	2.18	0.06	2.59	0.07	2.37	0.06	1.94	0.10	1.86	0.05	1.89	0.05
Large SUV/van	1.99	0.07	2.19	0.07	2.13	0.09	1.89	0.06	1.78	0.08	1.56	0.04
Small truck	1.91	0.06	3.01	0.06	2.85	0.07	2.14	0.06	2.00	0.06	1.92	0.07
Large truck	1.35	0.04	2.98	0.10	1.75	0.07	1.72	0.08	1.52	0.03	1.51	0.04
Minivan	2.75	0.09	2.95	0.08	3.11	0.10	2.92	0.11	2.00	0.04	2.04	0.07
λ parameter												
Constant	-4.33	0.17	-3.54	0.04	-3.26	0.06	-3.23	0.04	-3.42	0.05	-3.58	0.04
τ parameter												
Luxury compact	-6.81	0.20	-4.80	0.11	-4.86	0.09	-7.32	0.17	-7.19	0.17	-5.40	0.14
Midsize	-3.84	0.18	-1.22	0.14	-0.72	0.08	-2.90	0.14	-2.65	0.10	-2.56	0.14
Fullsize	-2.54	0.15	-2.83	0.24	-1.29	0.11	-4.86	0.16	-4.07	0.10	-2.54	0.18
Luxury midsize/fullsize	-4.54	0.22	-3.79	0.19	-2.37	0.12	-5.90	0.17	-5.45	0.11	-3.44	0.19
Small SUV	-6.49	0.14	-4.91	0.19	-3.56	0.10	-5.09	0.23	-5.36	0.11	-5.95	0.08
Large SUV/van	-2.62	0.19	-4.81	0.17	-4.56	0.21	-3.93	0.24	-4.29	0.10	-3.49	0.11
Small truck	-4.59	0.16	-5.41	0.09	-4.64	0.08	-4.58	0.19	-4.74	0.13	-4.74	0.22
Large truck	-0.52	0.12	-2.90	0.13	-2.38	0.17	-2.01	0.10	-1.11	0.11	-0.41	0.11
Minivan	-6.89	0.16	-5.82	0.19	-6.29	0.18	-7.34	0.10	-7.02	0.12	-6.15	0.11
Ford	-4.92	0.13	-4.09	0.13	-3.89	0.10	-5.27	0.23	-4.93	0.27	-4.36	0.16
Chrysler	-6.70	0.23	-5.76	0.16	-6.02	0.29	-6.48	0.14	-6.22	0.23	-5.81	0.12
GM	-4.50	0.25	-2.62	0.08	-2.67	0.12	-4.52	0.18	-3.65	0.18	-3.17	0.14
Honda	-4.37	0.08	-3.99	0.12	-6.24	0.07	-4.76	0.20	-6.31	0.23	-6.25	0.09
Toyota	-6.04	0.17	-4.24	0.14	-5.20	0.10	-7.38	0.15	-7.39	0.11	-5.82	0.21
Other East Asian	-6.19	0.25	-5.21	0.10	-6.09	0.13	-5.72	0.24	-6.53	0.12	-7.24	0.12
European	-7.52	0.16	-7.03	0.17	-7.87	0.17	-7.91	0.24	-9.47	0.14	-8.48	0.23
Age category #1 – 2001-02**	-3.99	0.10	-1.31	0.06	-3.67	0.09	-2.99	0.13	-2.45	0.15	-3.02	0.11
Age category #2 – 1999-00**	-2.16	0.10	-0.99	0.07	-3.57	0.09	-1.88	0.10	-1.16	0.13	-2.56	0.11
Age category #3 – 1995-98**	-4.18	0.11	-3.48	0.08	-3.50	0.16	-4.83	0.14	-5.04	0.12	-4.24	0.10
Age category #4 – 1990-94**	-5.08	0.12	-4.61	0.18	-4.16	0.12	-3.87	0.11	-3.91	0.16	-5.03	0.11
Weight / 100**	-0.14	0.00	-0.16	0.01	-0.19	0.00	-0.15	0.00	-0.13	0.00	-0.15	0.00
Wheelbase / 100**	-5.14	0.16	-6.43	0.12	-5.23	0.26	-5.49	0.19	-5.76	0.16	-5.42	0.12
HP / weight**	-0.21	0.01	-0.16	0.01	-0.16	0.01	-0.19	0.01	-0.18	0.01	-0.20	0.01
ϕ parameter												
MSA < 250k	0.06	0.18	0.24	0.17	0.13	0.09	0.33	0.14	1.30	0.10	-0.19	0.09
MSA < 500k & ≥ 250k	1.98	0.20	2.45	0.17	2.29	0.09	2.42	0.10	1.81	0.09	1.86	0.12
MSA < 1m & ≥ 500k	0.70	0.12	1.41	0.09	0.46	0.08	1.67	0.11	2.46	0.12	1.26	0.14
MSA < 3m & ≥ 1m	2.36	0.15	1.40	0.13	0.90	0.13	2.04	0.23	2.63	0.18	2.38	0.10

MSA ≥ 3m	4.14	0.14	2.94	0.18	2.54	0.08	3.29	0.13	3.79	0.16	4.64	0.13
White respondent	2.85	0.13	1.64	0.08	1.88	0.11	0.82	0.11	0.58	0.08	4.25	0.18
HS diplomas per adult	6.35	0.17	3.09	0.14	0.74	0.08	5.08	0.21	6.21	0.16	5.92	0.11
4-yr college deg. per adult	-0.62	0.13	-0.40	0.11	-1.30	0.06	-0.36	0.17	-0.42	0.12	-0.04	0.07
Average adult age	2.35	0.24	4.51	0.08	4.39	0.18	-0.38	0.24	1.66	0.23	1.77	0.15
Other parameters												
μ^*	1.15	0.02	0.77	0.02	0.64	0.02	0.98	0.02	1.18	0.02	1.34	0.02
σ^*	-0.04	0.08	-0.27	0.09	-0.77	0.06	1.07	0.05	0.70	0.05	0.44	0.06

All estimates generated with 40,000 iterations of the Gibbs sampling algorithm. The first 30,000 iterations were discarded as burn-in, and every 10th iteration thereafter was used to construct the reported estimates.

** For the tau parameter, weight wheelbase, and hp / weight are restricted to be positive. Also, the age dummies are restricted to be positive and monotonically increasing in the newness of the car. That is, the age coefficient for a car produced between 1990 and 1994 is $\exp(\beta_{\text{age category \#4}})$, whereas the age coefficient for a car produced in 2000 or 2001 is $\exp(\beta_{\text{age category \#4}} + \beta_{\text{age category \#3}} + \beta_{\text{age category \#2}} + \beta_{\text{age category \#1}})$.

Table 3a
Posterior Mean Parameter Estimates – Strata 7-12

Strata	*#7*		*#8*		*#9*		*#10*		*#11*		*#12*	
	Mean	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*
α parameter												
Age category #1 – 2001-02	2.17	0.06	2.32	0.07	1.81	0.07	1.68	0.06	1.44	0.07	1.35	0.18
Age category #2 – 1999-00	2.17	0.06	2.36	0.05	1.90	0.07	1.83	0.12	1.56	0.07	1.40	0.18
Age category #3 – 1995-98	2.24	0.06	2.45	0.05	2.00	0.05	2.00	0.11	1.72	0.05	1.49	0.16
Age category #4 – 1990-94	1.89	0.06	2.11	0.04	1.69	0.05	1.72	0.10	1.48	0.05	1.27	0.16
Age category #5 – 1983-89	1.23	0.07	1.45	0.05	0.94	0.09	1.15	0.11	0.93	0.05	0.79	0.15
Horsepower (HP) / weight	0.02	0.01	0.01	0.01	-0.01	0.00	0.01	0.00	0.02	0.00	0.02	0.01
(HP / weight)*(avg adult age)	-3.67	0.69	-4.93	0.98	-0.89	0.55	-8.58	0.53	-4.43	0.60	-2.32	0.47
# of females	-0.09	0.07	-0.29	0.08	0.34	0.08	0.26	0.19	0.19	0.10	-0.01	0.05
# of workers	-	-	-	-	0.68	0.05	0.44	0.08	0.64	0.09	0.56	0.11
# of kids ≤ 17	-	-	-	-	0.46	0.06	-0.06	0.07	0.60	0.09	0.48	0.12
# of kids ≤ 11	-	-	-	-	0.09	0.07	0.56	0.09	-0.24	0.08	-	-
# of kids ≤ 6	-	-	-	-	-0.26	0.07	-0.42	0.16	-	-	-	-
# of kids ≤ 2	-	-	-	-	-0.31	0.08	-	-	-	-	-	-
β parameter												
Compact	2.01	0.04	1.92	0.06	1.82	0.03	1.86	0.03	1.99	0.03	1.89	0.04
Luxury compact	2.47	0.07	2.48	0.09	2.88	0.07	2.82	0.08	3.19	0.09	2.99	0.07
Midsize	1.89	0.06	1.45	0.05	1.91	0.04	1.98	0.05	2.05	0.05	1.98	0.05
Fullsize	1.85	0.04	1.36	0.04	2.06	0.11	2.17	0.05	2.25	0.08	2.05	0.06
Luxury midsize/fullsize	2.00	0.06	1.33	0.03	2.71	0.07	2.74	0.14	2.77	0.11	2.56	0.06
Small SUV	2.07	0.08	2.50	0.09	1.93	0.04	2.17	0.06	2.33	0.05	2.02	0.04
Large SUV/van	1.73	0.04	1.98	0.06	1.70	0.04	1.70	0.06	1.68	0.05	1.60	0.04
Small truck	2.18	0.03	2.50	0.09	2.41	0.05	2.30	0.05	2.30	0.05	2.34	0.06
Large truck	1.64	0.03	1.71	0.07	1.70	0.05	1.65	0.05	1.65	0.05	1.66	0.06
Minivan	2.03	0.05	1.84	0.08	1.79	0.05	1.85	0.04	1.87	0.03	1.85	0.08
λ parameter												
Constant	-3.31	0.05	-3.66	0.13	-3.13	0.06	-3.02	0.06	-3.26	0.04	-3.24	0.04
τ parameter												
Luxury compact	-3.98	0.20	-7.31	0.11	-5.23	0.20	-5.94	0.18	-4.24	0.29	-3.65	0.16
Midsize	-0.93	0.15	-1.51	0.14	-1.61	0.13	-1.87	0.11	-2.01	0.14	-1.97	0.14
Fullsize	-0.84	0.11	0.07	0.16	-2.73	0.12	-2.41	0.09	-2.65	0.17	-2.39	0.12
Luxury midsize/fullsize	-1.67	0.12	-3.05	0.28	-4.08	0.29	-3.53	0.18	-4.13	0.32	-3.28	0.10
Small SUV	-4.78	0.07	-7.18	0.29	-4.65	0.25	-3.06	0.14	-3.22	0.17	-3.02	0.10
Large SUV/van	-2.85	0.12	-4.06	0.08	-1.70	0.16	-1.83	0.19	-1.81	0.16	-2.16	0.11
Small truck	-2.82	0.11	-4.71	0.14	-3.21	0.30	-2.72	0.08	-3.37	0.25	-2.93	0.15
Large truck	-0.26	0.07	-1.93	0.11	-1.45	0.14	-1.02	0.10	-1.40	0.08	-0.83	0.12
Minivan	-3.26	0.11	-5.38	0.10	-2.01	0.21	-1.61	0.09	-1.66	0.10	-2.91	0.11
Ford	-1.96	0.19	-3.95	0.12	-2.93	0.18	-3.36	0.10	-2.77	0.16	-1.28	0.13
Chrysler	-4.03	0.12	-6.33	0.20	-3.67	0.16	-4.03	0.08	-3.46	0.11	-2.37	0.12
GM	-0.72	0.12	-1.82	0.11	-2.46	0.15	-2.88	0.10	-2.55	0.11	-0.94	0.17
Honda	-3.71	0.15	-6.90	0.16	-3.80	0.15	-4.90	0.12	-4.57	0.21	-2.93	0.18
Toyota	-3.94	0.11	-7.49	0.25	-4.94	0.13	-5.44	0.12	-4.99	0.10	-3.03	0.15
Other East Asian	-4.85	0.15	-7.68	0.21	-4.10	0.13	-4.97	0.18	-4.69	0.14	-3.04	0.17
European	-5.39	0.16	-8.21	0.07	-6.51	0.20	-6.75	0.15	-6.40	0.43	-5.79	0.18
Age category #1 – 2001-02**	-2.02	0.10	-2.90	0.16	-1.63	0.07	-2.51	0.14	-1.97	0.07	-1.94	0.19
Age category #2 – 1999-00**	-1.69	0.15	-2.31	0.13	-2.56	0.14	-1.39	0.10	-2.03	0.13	-1.46	0.15
Age category #3 – 1995-98**	-2.40	0.15	-3.23	0.08	-2.50	0.10	-3.04	0.09	-2.16	0.14	-2.71	0.09
Age category #4 – 1990-94**	-1.61	0.08	-3.47	0.10	-2.89	0.06	-3.14	0.10	-1.92	0.12	-3.56	0.11
Weight / 100**	-0.08	0.00	-0.12	0.00	-0.08	0.00	-0.08	0.00	-0.08	0.00	-0.09	0.00
Wheelbase / 100**	-3.01	0.06	-4.50	0.13	-3.68	0.15	-3.64	0.10	-3.45	0.08	-3.47	0.14
HP / weight**	-0.16	0.00	-0.28	0.01	-0.20	0.01	-0.15	0.01	-0.18	0.01	-0.17	0.01
ϕ parameter												

MSA < 250k	-0.45	0.11	-0.51	0.09	0.94	0.12	1.36	0.15	1.12	0.11	1.09	0.23
MSA < 500k & ≥ 250k	1.29	0.10	2.45	0.08	1.25	0.07	1.26	0.28	1.11	0.29	0.51	0.24
MSA < 1m & ≥ 500k	0.44	0.10	0.48	0.17	1.20	0.13	1.74	0.13	1.69	0.09	1.77	0.23
MSA < 3m & ≥ 1m	1.72	0.06	0.98	0.10	1.62	0.08	2.37	0.13	2.12	0.13	1.77	0.09
MSA ≥ 3m	2.28	0.07	1.61	0.06	2.72	0.08	2.81	0.11	3.03	0.08	2.48	0.13
White respondent	3.08	0.14	3.52	0.07	1.58	0.13	1.81	0.07	1.41	0.08	1.53	0.15
HS diplomas per adult	3.15	0.12	2.95	0.22	3.67	0.18	2.04	0.09	2.60	0.15	4.09	0.15
4-yr college deg. per adult	-1.59	0.17	-0.80	0.09	-0.86	0.16	-0.86	0.16	0.04	0.08	-1.10	0.10
Average adult age	5.80	0.31	5.73	0.43	4.98	0.37	3.97	0.26	3.44	0.19	3.60	0.16
Other parameters												
μ^*	0.85	0.02	1.05	0.02	0.82	0.03	0.74	0.02	0.81	0.04	0.80	0.03
σ^*	0.41	0.05	-0.25	0.07	0.81	0.07	0.84	0.04	0.77	0.05	1.06	0.06

All estimates generated with 40,000 iterations of the Gibbs sampling algorithm. The first 30,000 iterations were discarded as burn-in, and every 10th iteration thereafter was used to construct the reported estimates.

** For the tau parameter, weight wheelbase, and hp / weight are restricted to be positive. Also, the age dummies are restricted to be positive and monotonically increasing in the newness of the car. That is, the age coefficient for a car produced between 1990 and 1994 is $\exp(\beta_{\text{age category \#4}})$, whereas the age coefficient for a car produced in 2000 or 2001 is $\exp(\beta_{\text{age category \#4}} + \beta_{\text{age category \#3}} + \beta_{\text{age category \#2}} + \beta_{\text{age category \#1}})$.

Table 4a
Posterior Variance Parameter Estimates – Strata 1-6

Strata	***#1***		***#2***		***#3***		***#4***		***#5***		***#6***	
	Mean	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*
α parameter												
Age category #1 – 2001-02	1.44	0.13	0.94	0.08	1.05	0.11	0.79	0.08	0.90	0.11	0.86	0.06
Age category #2 – 1999-00	1.22	0.11	0.87	0.13	0.95	0.10	0.71	0.09	0.71	0.10	0.76	0.05
Age category #3 – 1995-98	0.96	0.08	0.80	0.10	0.87	0.07	0.62	0.06	0.75	0.09	0.73	0.05
Age category #4 – 1990-94	0.94	0.07	0.75	0.09	0.88	0.06	0.64	0.07	0.71	0.08	0.71	0.05
Age category #5 – 1983-89	0.98	0.09	0.99	0.08	1.07	0.11	0.94	0.11	0.79	0.09	1.00	0.11
Horsepower (HP) / weight	0.07	0.01	0.04	0.00	0.06	0.01	0.09	0.01	0.06	0.01	0.05	0.01
(HP / weight)*(avg adult age)	11.64	1.42	8.66	1.57	6.96	0.55	16.12	1.52	12.83	1.08	10.92	0.95
# of females	-	-	-	-	0.43	0.03	0.83	0.10	1.02	0.10	1.10	0.11
# of workers	-	-	-	-	-	-	0.78	0.12	0.57	0.07	0.56	0.05
β parameter												
Compact	0.74	0.08	0.53	0.04	0.75	0.07	0.60	0.05	0.55	0.03	0.67	0.04
Luxury compact	2.37	0.20	1.50	0.14	1.97	0.21	1.43	0.14	1.56	0.10	0.97	0.08
Midsize	0.75	0.07	0.64	0.06	0.62	0.05	0.69	0.06	0.61	0.06	0.64	0.04
Fullsize	1.13	0.12	0.92	0.10	0.62	0.04	0.89	0.13	0.60	0.06	0.71	0.06
Luxury midsize/fullsize	1.66	0.19	1.22	0.12	1.04	0.06	1.38	0.22	0.91	0.09	0.83	0.06
Small SUV	1.38	0.13	1.42	0.17	1.38	0.11	0.75	0.12	0.72	0.05	0.94	0.05
Large SUV/van	0.97	0.07	1.22	0.12	1.33	0.11	0.65	0.06	0.60	0.10	0.61	0.05
Small truck	0.92	0.10	1.45	0.12	1.48	0.16	0.71	0.06	0.66	0.08	0.66	0.05
Large truck	0.85	0.10	1.55	0.12	0.84	0.12	0.65	0.07	0.49	0.05	0.56	0.04
Minivan	1.75	0.14	1.49	0.13	2.40	0.22	1.18	0.11	0.75	0.06	0.90	0.07
λ parameter												
Constant	1.27	0.19	0.77	0.09	1.06	0.09	0.70	0.09	0.59	0.05	0.52	0.05
τ parameter												
Luxury compact	7.80	0.94	3.22	0.50	2.76	0.26	3.92	0.65	13.58	1.16	4.10	0.51
Midsize	12.37	1.48	3.85	0.41	4.69	0.57	4.62	0.48	6.21	0.50	8.87	0.97
Fullsize	4.17	0.60	6.18	0.86	5.18	0.39	3.88	0.53	5.85	0.47	6.96	0.74
Luxury midsize/fullsize	2.11	0.32	4.36	0.74	3.20	0.23	2.02	0.31	3.78	0.30	5.36	1.03
Small SUV	5.67	0.54	7.85	1.02	1.48	0.15	11.00	0.98	6.00	0.59	3.25	0.41
Large SUV/van	3.92	0.77	4.62	0.49	3.41	0.24	3.90	0.41	7.20	0.60	5.78	0.48
Small truck	9.77	1.69	4.65	0.38	3.64	0.51	5.40	1.11	6.15	1.22	7.56	0.79
Large truck	4.93	0.47	2.73	0.29	1.88	0.21	2.54	0.34	2.81	0.30	3.10	0.21
Minivan	3.47	0.34	3.53	0.30	3.13	0.39	3.82	0.54	5.06	0.54	3.16	0.44
Ford	5.36	0.60	6.48	0.58	4.84	0.51	7.39	0.62	15.27	1.91	17.69	1.76
Chrysler	9.50	1.01	7.22	0.77	5.83	0.99	4.55	0.63	9.73	0.78	8.02	0.83
GM	27.23	2.83	4.57	0.54	5.82	0.58	10.15	1.34	10.72	1.20	14.73	1.75
Honda	2.13	0.38	2.10	0.23	2.50	0.35	3.26	0.41	5.33	0.98	5.78	0.49
Toyota	9.23	1.33	3.96	0.36	2.96	0.38	8.78	0.89	5.51	0.54	6.04	0.77
Other East Asian	10.74	0.97	6.07	0.53	2.53	0.17	5.01	1.16	8.19	1.11	11.17	1.23
European	4.16	0.55	3.18	0.41	3.34	0.59	7.91	0.64	8.21	1.12	7.56	0.67
Age category #1 – 2001-02**	1.99	0.21	1.23	0.10	2.08	0.25	2.84	0.36	2.30	0.19	2.60	0.24
Age category #2 – 1999-00**	1.85	0.30	1.18	0.11	2.62	0.30	2.73	0.39	1.82	0.18	3.42	0.26
Age category #3 – 1995-98**	2.61	0.23	2.18	0.27	2.88	0.30	2.95	0.33	3.03	0.55	2.06	0.23
Age category #4 – 1990-94**	4.27	0.41	2.02	0.19	2.61	0.24	3.01	0.31	3.00	0.27	4.07	0.60
Weight / 100**	0.14	0.01	0.08	0.01	0.09	0.02	0.11	0.01	0.07	0.01	0.12	0.02
Wheelbase / 100**	3.83	0.52	2.54	0.50	2.29	0.39	3.67	0.49	2.44	0.28	2.92	0.24
HP / weight**	0.13	0.02	0.11	0.02	0.10	0.01	0.10	0.02	0.13	0.02	0.15	0.01
ϕ parameter												
MSA < 250k	2.01	0.27	4.37	0.43	2.22	0.20	2.95	0.33	3.90	0.33	2.29	0.23
MSA < 500k & ≥ 250k	4.03	0.39	2.33	0.56	3.48	0.48	3.69	0.36	2.00	0.31	2.42	0.23
MSA < 1m & ≥ 500k	3.82	0.35	2.58	0.32	2.16	0.42	2.24	0.49	3.79	0.52	5.18	0.60
MSA < 3m & ≥ 1m	3.52	0.64	1.84	0.24	1.89	0.27	4.76	0.49	2.50	0.47	2.30	0.26

MSA ≥ 3m	4.13	0.38	3.66	0.47	2.20	0.22	2.06	0.19	3.55	0.42	3.82	0.67
White respondent	4.73	0.90	2.26	0.41	3.52	0.45	1.80	0.38	2.39	0.21	2.93	0.39
HS diplomas per adult	4.65	1.09	2.72	0.21	2.09	0.18	3.58	0.69	4.19	0.46	3.38	0.34
4-yr college deg. per adult	2.61	0.43	2.19	0.33	1.89	0.17	2.00	0.31	3.55	0.36	3.41	0.43
Average adult age	10.55	1.07	2.76	0.36	4.75	0.82	6.52	1.66	3.52	0.36	4.30	0.48
Other parameters												
μ^*	0.23	0.02	0.19	0.01	0.17	0.01	0.19	0.01	0.17	0.01	0.16	0.01
σ^*	1.59	0.14	1.27	0.10	1.10	0.08	0.78	0.05	1.00	0.06	1.12	0.06

All estimates generated with 40,000 iterations of the Gibbs sampling algorithm. The first 30,000 iterations were discarded as burn-in, and every 10th iteration thereafter was used to construct the reported estimates.

** For the tau parameter, weight wheelbase, and hp / weight are restricted to be positive. Also, the age dummies are restricted to be positive and monotonically increasing in the newness of the car. That is, the age coefficient for a car produced between 1990 and 1994 is $\exp(\beta_{\text{age category \#4}})$, whereas the age coefficient for a car produced in 2000 or 2001 is $\exp(\beta_{\text{age category \#4}} + \beta_{\text{age category \#3}} + \beta_{\text{age category \#2}} + \beta_{\text{age category \#1}})$.

Table 5a
Posterior Variance Parameter Estimates – Strata 7-12

Strata	***#7***		***#8***		***#9***		***#10***		***#11***		***#12***	
	Mean	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*	*Mean*	*St.Er.*
α parameter												
Age category #1 – 2001-02	0.86	0.07	1.06	0.13	0.85	0.09	0.93	0.13	0.82	0.09	0.86	0.12
Age category #2 – 1999-00	0.79	0.08	0.79	0.07	0.64	0.06	0.74	0.07	0.78	0.06	0.73	0.09
Age category #3 – 1995-98	0.63	0.06	0.71	0.07	0.60	0.05	0.60	0.04	0.66	0.05	0.61	0.05
Age category #4 – 1990-94	0.71	0.06	0.74	0.07	0.69	0.07	0.57	0.04	0.73	0.06	0.67	0.06
Age category #5 – 1983-89	0.79	0.07	0.89	0.07	0.87	0.09	0.81	0.08	0.85	0.07	0.80	0.11
Horsepower (HP) / weight	0.05	0.01	0.08	0.01	0.05	0.01	0.07	0.01	0.08	0.01	0.05	0.01
(HP / weight)*(avg adult age)	7.48	0.66	12.08	1.12	19.35	3.29	15.26	1.49	8.58	0.96	9.10	1.05
# of females	0.87	0.06	1.73	0.19	1.81	0.23	1.45	0.14	1.24	0.14	1.08	0.17
# of workers	-	-	-	-	0.81	0.07	1.07	0.12	1.01	0.12	0.78	0.09
# of kids ≤ 17	-	-	-	-	1.40	0.21	0.93	0.08	1.26	0.12	1.56	0.21
# of kids ≤ 11	-	-	-	-	1.66	0.19	1.25	0.20	1.31	0.14	-	-
# of kids ≤ 6	-	-	-	-	1.63	0.16	1.39	0.25	-	-	-	-
# of kids ≤ 2	-	-	-	-	1.86	0.24	-	-	-	-	-	-
β parameter												
Compact	0.60	0.04	0.64	0.08	0.48	0.04	0.54	0.05	0.46	0.03	0.49	0.04
Luxury compact	0.94	0.07	1.39	0.20	1.18	0.15	1.48	0.11	1.51	0.19	1.52	0.12
Midsize	0.62	0.04	0.67	0.05	0.54	0.05	0.60	0.05	0.61	0.04	0.49	0.05
Fullsize	0.50	0.06	0.56	0.03	0.70	0.10	0.72	0.07	0.73	0.05	0.55	0.05
Luxury midsize/fullsize	0.68	0.07	0.71	0.04	1.21	0.09	1.26	0.11	1.21	0.25	0.94	0.08
Small SUV	0.84	0.09	1.38	0.13	0.70	0.05	0.79	0.09	0.81	0.06	0.70	0.06
Large SUV/van	0.52	0.04	0.89	0.06	0.52	0.04	0.53	0.05	0.44	0.04	0.50	0.04
Small truck	0.70	0.05	1.03	0.11	0.81	0.05	0.72	0.07	0.74	0.07	0.72	0.08
Large truck	0.50	0.03	0.62	0.05	0.56	0.04	0.53	0.05	0.47	0.05	0.42	0.03
Minivan	0.65	0.05	1.12	0.09	0.53	0.03	0.57	0.06	0.52	0.06	0.55	0.05
λ parameter												
Constant	0.63	0.04	0.88	0.12	0.50	0.04	0.51	0.04	0.52	0.04	0.54	0.04
τ parameter												
Luxury compact	5.41	0.69	3.68	0.74	3.63	0.39	5.83	1.04	3.00	0.51	2.37	0.41
Midsize	3.09	0.35	9.21	1.52	2.92	0.30	2.48	0.27	2.57	0.42	3.38	0.84
Fullsize	3.90	0.49	3.53	0.29	2.34	0.32	4.20	0.42	3.65	0.34	3.26	0.30
Luxury midsize/fullsize	1.76	0.21	5.27	0.92	1.91	0.25	3.49	0.88	4.26	0.87	4.19	0.62
Small SUV	3.13	0.29	4.46	0.68	7.85	1.05	3.30	0.23	2.55	0.43	3.70	0.48
Large SUV/van	5.48	0.54	3.40	0.37	3.74	0.41	4.04	0.57	4.20	0.64	3.82	0.79
Small truck	5.10	0.60	7.55	0.67	3.15	0.44	1.89	0.17	2.79	0.35	3.38	0.36
Large truck	1.78	0.27	3.61	0.34	2.34	0.19	2.35	0.44	2.58	0.22	2.53	0.32
Minivan	3.51	0.50	5.87	1.05	2.98	0.32	2.75	0.57	3.36	0.39	3.72	0.39
Ford	8.88	1.08	9.72	0.85	7.04	0.85	3.36	0.43	6.88	0.59	10.04	1.39
Chrysler	7.01	0.50	6.97	1.20	2.97	0.53	3.65	0.36	5.47	0.97	5.77	0.67
GM	8.72	0.93	6.90	0.55	4.67	0.36	2.38	0.22	6.21	0.68	8.90	0.97
Honda	4.27	0.32	2.50	0.27	4.44	0.53	4.60	0.50	2.77	0.25	2.34	0.49
Toyota	5.15	0.50	10.60	1.92	4.56	0.66	5.22	0.47	2.95	0.32	2.54	0.23
Other East Asian	4.67	0.46	2.10	0.21	2.58	0.45	6.96	0.85	3.04	0.34	3.10	0.37
European	2.96	0.52	2.20	0.54	3.83	0.40	4.76	0.68	3.81	0.66	9.88	1.45
Age category #1 – 2001-02**	1.66	0.12	2.49	0.29	1.45	0.19	1.95	0.19	1.51	0.11	2.26	0.29
Age category #2 – 1999-00**	1.76	0.24	1.71	0.19	3.14	0.28	1.37	0.13	1.81	0.16	1.78	0.24
Age category #3 – 1995-98**	1.71	0.22	2.66	0.39	1.71	0.17	2.02	0.20	1.48	0.14	1.31	0.23
Age category #4 – 1990-94**	1.40	0.20	3.24	0.28	1.84	0.17	1.78	0.22	1.51	0.12	2.46	0.38
Weight / 100**	0.05	0.01	0.07	0.01	0.04	0.00	0.05	0.00	0.04	0.00	0.06	0.01
Wheelbase / 100**	1.74	0.13	2.45	0.26	2.00	0.28	2.40	0.32	1.60	0.23	1.79	0.26
HP / weight**	0.10	0.01	0.11	0.01	0.10	0.01	0.09	0.01	0.09	0.01	0.09	0.02
ϕ parameter												

MSA < 250k	2.86	0.48	2.98	0.20	2.43	0.21	2.67	0.18	2.13	0.31	3.11	0.29
MSA < 500k & ≥ 250k	1.67	0.19	1.85	0.19	2.28	0.27	2.70	0.67	2.56	0.30	4.34	1.15
MSA < 1m & ≥ 500k	2.30	0.48	2.66	0.27	2.76	0.28	2.33	0.22	1.63	0.17	2.15	0.26
MSA < 3m & ≥ 1m	1.91	0.18	1.92	0.17	1.72	0.26	1.52	0.31	1.81	0.21	2.18	0.24
MSA ≥ 3m	2.19	0.25	2.18	0.28	1.85	0.18	1.94	0.17	1.81	0.21	1.84	0.21
White respondent	6.90	0.89	1.91	0.21	2.47	0.34	1.78	0.21	3.86	0.38	2.99	0.47
HS diplomas per adult	3.08	0.32	3.69	0.42	5.97	0.63	2.55	0.28	2.61	0.24	4.88	0.55
4-yr college deg. per adult	2.01	0.17	1.91	0.42	1.62	0.11	1.78	0.22	1.61	0.23	2.76	0.28
Average adult age	8.46	0.97	4.36	0.44	3.67	0.61	3.23	0.70	2.41	0.24	3.62	0.39
Other parameters												
μ^*	0.15	0.01	0.14	0.01	0.18	0.01	0.18	0.01	0.18	0.01	0.21	0.02
σ^*	0.92	0.09	1.16	0.07	0.96	0.06	0.91	0.05	0.91	0.08	0.79	0.05

All estimates generated with 40,000 iterations of the Gibbs sampling algorithm. The first 30,000 iterations were discarded as burn-in, and every 10th iteration thereafter was used to construct the reported estimates.

** For the tau parameter, weight wheelbase, and hp / weight are restricted to be positive. Also, the age dummies are restricted to be positive and monotonically increasing in the newness of the car. That is, the age coefficient for a car produced between 1990 and 1994 is $\exp(\beta_{\text{age category \#4}})$, whereas the age coefficient for a car produced in 2000 or 2001 is $\exp(\beta_{\text{age category \#4}} + \beta_{\text{age category \#3}} + \beta_{\text{age category \#2}} + \beta_{\text{age category \#1}})$.

C Calibration of the Simulation Model

The individual household utility functions (and the associated system of automobile and VMT demands) are characterized entirely using the parameter estimates described in section 4 of the main text. The simulation model, however, also generates a supply function for both new and used cars that needs to be calibrated separately. The following two subsections describe the calibration method for the new and used car supply functions, respectively.

1 Calibration of new car supply

The producer problem described in section 2-c of the paper requires as input the marginal costs faced by each producer for each class of vehicle. We calibrate these costs using the dealer markups available on www.edmunds.com[7] and the estimated ratio of dealer and manufacturer markups from Bresnahan (1986). This provides the following estimated total markup by class and manufacturer:

Table A-1: Markups by Manufacturer and Class

Class:	Ford	Chrysler	GM	Honda	Toyota	Other Asian	European
Compact	15	14	22	28	22	23	23
Lux compact		18	22	27	41	25	27
Midsize	19	21	24	34	33	16	26
Fullsize	16	20	26		38		
Lux mid/full	19	20	24	38	46		20
Small SUV	19	16	24	24	25	18	
Large SUV	37	27	41	33	43	23	29
Small truck	19	27	28		29	21	
Large truck	36	39	41		30		
Minivan	22	23	41	34	32	21	22

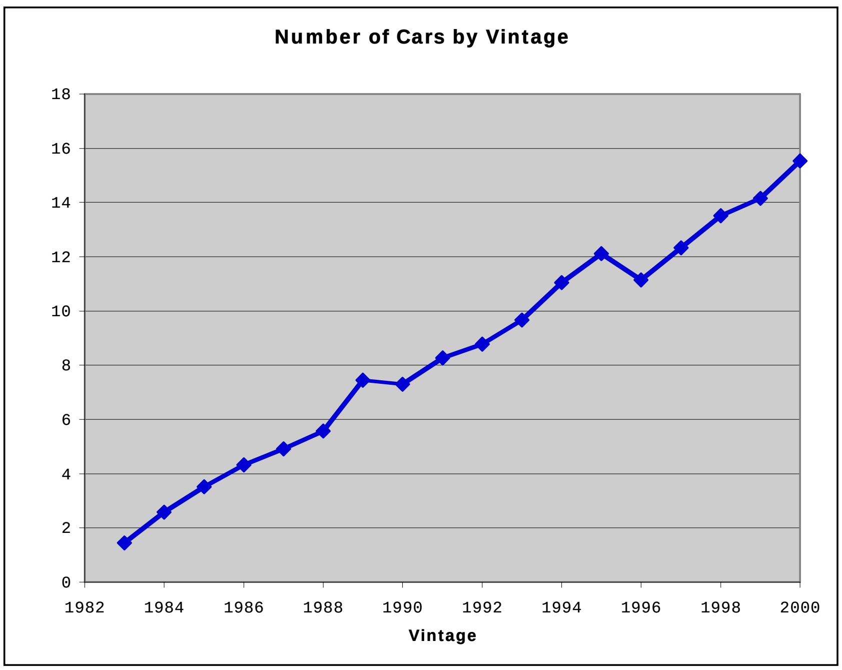
2 Calibration of used car supply (the scrap market)

The total quantity of used cars supplied of a given make, age, and class depends on how many are scrapped. The function controlling scrap is given in section 2-d-ii of the paper and

[7] www.edmunds.com provides the invoice prices and suggested retail prices for automobiles by make and model. We use data for 2001 (corresponding to our household sample period).

requires both a calibration constant determining scrap level in the baseline case and an elasticity controlling how rapidly the quantity of scrapped vehicles changes. The maximum possible supply of a particular used vehicle is determined by how many were available in the market in the previous time period.

We use the parameter b_j to calibrate the probability a vehicle is scrapped in the benchmark. The calibration is based on the roughly linear trend of vehicle choice that can be observed in the data. The figure below shows the number of used cars of each vintage in our NHTS dataset:



We take the quantity of vehicles scrapped of each vintage in each year to be constant (as suggested by the figure), implying that the scrap rate for a vehicle of a given age is simply:

$$\theta_{age} = \frac{1}{20 - age} \quad \text{(for } age \text{ between 1 and 19)}$$

Note that the scrap rate for cars beginning their 19th year is 100%, meaning none enter the used car market in our simulations.

After establishing the benchmark progression of scrap rates, we also calibrate the response of the scrap market to changes in vehicle values: The parameter controlling this response is the elasticity η_j. We take the aggregate results from Alberini et al. (1998) who find that a $1000 bounty (equivalent to 67% of the average vehicle value) causes a 193% increase in the number of vehicles scrapped. This implies an elasticity of 2.9. We adopt the round figure of 3 for our central case and double the value to 6 in our sensitivity analysis.

Appendix References:

Alberini, A., Harrington, W., and V. McConnell (1998), "Fleet Turnover and Old Car Scrap Policies," *RFF Discussion Paper* 98-23, March 1998.

Allenby, G., and P. Lenk (1994), "Modeling Household Purchase Behavior with Logistic Normal Regression," *Journal of the American Statistical Association*, 89, 669-679.

Bresnahan, T. (1986), "Departures from Marginal-Cost Pricing in the American Automobile Industry: Estimates for 1977-1978," *Journal of Econometrics*, 11, 201-227.

Chib, S., and E. Greenberg (1995), "Understanding the Metropolis-Hastings Algorithm," *American Statistician*, 49, 327-335.

Gelman, A, Carlin, J., Stern, H., and D. Rubin (1995), *Bayesian Data Analysis.* Chapman & Hall, London, UK.

Train, K.E. (2003), *Discrete Choice Analysis with Simulation.* Cambridge University Press, Cambridge, UK.

Table 2-1: Included Car Types

Classes	Age categories	Manufacturers
Compact	New cars	Ford
Luxury compact	1-2 years old	Chrysler
Midsize	3-6 years old	General Motors
Fullsize	7-11 years old	Honda
Luxury mid/fullsize	12-18 years old	Toyota
Small SUV		Other Asian
Large SUV		European
Small truck		
Large truck		
Minivan		

Table 3-1
Sample Demographic Statistics from the 2001 NHTS – 20,429 Observations

Variable	***mean (std. dev.)***
Household size	2.490 (1.34)
# of adults ≥ 18 years old	1.861 (0.69)
# of adults ≥ 65 years old	0.380 (0.67)
# of children ≤ 2 years old	0.096 (0.32)
# of children 3-6 years old	0.136 (0.41)
# of children 7-11 years old	0.185 (0.49)
# of children 12-17 years old	0.211 (0.54)
# of workers	1.272 (0.95)
# of females	1.033 (0.52)
Average age among adults (≥ 18)	49.56 (16.8)
Household income (2001 $s)	56,621 (43,276)
Household breakdown:	***percentage***
1 male adult, no children, not retired	5.71
1 female adult, no children, not retired	7.88
1 adult, no children, retired	10.3
2+ adults w/ average age ≤ 35, no children, not retired	7.10
2+ adults w/ average age > 35 & ≤ 50, no children, not retired	8.43
2+ adults w/ average age > 50, no children, not retired	9.04
2+ adults w/ average age ≤ 67, no children, retired	9.29
2+ adults w/ average age > 67, no children, retired	8.47
1+ adults w/ youngest child < 3 years old	8.69
1+ adults w/ youngest child 3-6 years old	7.65
1+ adults w/ youngest child 7-11 years old	8.64
1+ adults w/ youngest child 12-17 years old	8.85
White household respondent [b]	85.6
Black household respondent	7.62
Hispanic household respondent	6.25
Asian household respondent	2.17
Adults with high school diplomas	89.4
Adults with 4-year college degrees	30.5
Resident of MSA < 250k	7.62
Resident of MSA 250-500k	8.22
Resident of MSA 500k-1m	8.30
Resident of MSA 1-3m	22.2
Resident of MSA > 3m	32.5
Non-resident of MSA	21.1
Household income ≤$25,000	22.8
Household income ≤$50,000 & >$25,000	33.3
Household income ≤$75,000 & >$50,000	19.8
Household income >$75,000	24.1

[a] Based on 8 level index where 1=less than high school graduate, 2= high school graduate including GED, 3= vocational/technical training, 4=some college, but no degree, 5=Associate's degree, 6= Bachelor's degree, 7= some graduate or professional school, but no degree, 8= Graduate or professional school degree.
[b] The white, black, Hispanic, and Asian percentages sum to more than 100 percent because some respondents have multi-cultural backgrounds.

Table 3-2
Automobile Characteristics [a]

Characteristic	*Compact*	*Luxury Compact*	*Midsize*	*Fullsize*	*Luxury Mid/Full*	*Small SUV*	*Large SUV / Van*	*Small truck*	*Large truck*	*Minivan*	*Total*
Miles per Gallon[b]											
All Car Ages:	29.73 (1.55)	24.18 (1.29)	27.16 (1.72)	25.57 (2.15)	23.65 (0.85)	23.75 (2.28)	20.04 (2.96)	23.60 (2.35)	19.82 (2.10)	23.45 (1.68)	24.39 (3.46)
Model Years:											
2001-2002	30.29 (1.67)	24.47 (1.47)	26.90 (2.05)	25.61 (2.29)	23.70 (0.45)	24.17 (1.92)	19.08 (1.79)	22.62 (2.43)	19.65 (2.91)	23.25 (1.48)	24.15 (3.66)
1999-2000	30.32 (2.47)	24.45 (1.26)	27.29 (1.41)	25.79 (2.12)	23.86 (0.50)	23.80 (2.60)	18.21 (1.20)	23.29 (2.89)	20.34 (2.12)	22.99 (1.80)	24.18 (3.80)
1995-1998	30.02 (1.23)	24.24 (1.37)	27.50 (1.47)	25.51 (1.80)	24.29 (0.74)	23.44 (2.63)	19.60 (2.74)	23.30 (2.59)	18.67 (0.85)	23.65 (1.97)	24.44 (3.55)
1990-1994	29.21 (0.90)	23.81 (1.42)	26.74 (1.55)	25.37 (2.42)	22.91 (1.15)	22.67 (2.72)	20.90 (3.84)	23.10 (1.67)	18.56 (1.28)	23.57 (1.36)	24.08 (3.35)
1983-1989	28.82 (0.46)	23.94 (1.25)	27.38 (2.35)	25.56 (3.31)	23.23 (0.90)	24.84 (1.36)	22.88 (3.11)	25.70 (1.44)	21.76 (1.69)	23.97 (2.16)	25.14 (2.78)
Horsepower/100											
All Car Ages:	1.286 (0.24)	2.275 (0.58)	1.530 (0.25)	1.726 (0.40)	2.177 (0.35)	1.531 (0.23)	1.909 (0.49)	1.386 (0.25)	2.011 (0.41)	1.575 (0.30)	1.719 (48.3)
Model Years:											
2001-2002	1.526 (0.14)	2.621 (0.69)	1.787 (0.13)	2.123 (0.11)	2.463 (0.27)	1.763 (0.11)	2.391 (0.13)	1.650 (0.09)	2.377 (0.28)	1.833 (0.27)	2.036 (0.46)
1999-2000	1.454 (0.14)	2.488 (0.60)	1.682 (0.08)	1.917 (0.24)	2.376 (0.20)	1.648 (0.15)	2.271 (0.14)	1.556 (0.14)	2.323 (0.22)	1.765 (0.23)	1.932 (0.43)
1995-1998	1.342 (0.13)	2.414 (0.61)	1.597 (0.09)	1.835 (0.26)	2.237 (0.20)	1.554 (0.17)	2.024 (0.12)	1.430 (0.15)	1.992 (0.10)	1.513 (0.22)	1.773 (0.42)
1990-1994	1.152 (0.06)	2.075 (0.35)	1.418 (0.10)	1.469 (0.39)	1.952 (0.12)	1.467 (0.12)	1.476 (0.35)	1.257 (0.18)	1.698 (0.12)	1.378 (0.15)	1.516 (0.35)
1983-1989	0.955 (0.06)	1.777 (0.26)	1.166 (0.16)	1.212 (0.24)	1.637 (0.26)	1.164 (0.10)	1.244 (0.21)	1.038 (0.07)	1.435 (0.11)	1.272 (0.22)	1.270 (0.30)

[a] Standard deviations reported in parentheses.
[b] Weighted harmonic mean of EPA test miles per gallon estimates.

Table 4-1: Posterior Mean Elasticity Estimates

	VMT elasticity wrt operating cost[1]	VMT elasticity wrt income[1]	Car ownership elasticity wrt rental price
All	-0.69	0.62	-0.82
By Household			
Retired	-0.64	0.53	-0.93
Not retired, no children	-0.64	0.54	-0.72
Not retired, with children	-0.76	0.77	-0.85
By Auto			
By Class			
All Cars			
Compact	-0.62	0.58	-0.65
Luxury compact	-1.10	0.77	-1.25
Midsize	-0.61	0.56	-0.67
Fullsize	-0.70	0.57	-0.73
Luxury midsize/fullsize	-0.82	0.68	-1.25
Small SUV	-0.66	0.67	-0.73
Large SUV/van	-0.72	0.77	-0.98
Small truck	-0.84	0.60	-0.62
Large truck	-0.72	0.63	-0.85
Minivan	-0.71	0.67	-0.77
New Cars			
Compact	-0.52	0.76	-1.44
Luxury compact	-0.65	1.18	-3.14
Midsize	-0.52	0.81	-1.58
Fullsize	-0.54	0.81	-1.77
Luxury midsize/fullsize	-0.46	1.30	-3.04
Small SUV	-0.52	0.82	-1.58
Large SUV/van	-0.56	1.10	-2.30
Small truck	-0.67	0.83	-1.32
Large truck	-0.59	0.83	-1.69
Minivan	-0.58	0.95	-1.67
By Age			
New cars	-0.55	0.88	-1.97
1-2 year old cars	-0.61	0.73	-1.01
3-6 year old cars	-0.66	0.62	-0.73
7-11 year old cars	-0.75	0.54	-0.28
12-18 year old cars	-0.83	0.50	-0.13

[1] All VMT elasticities are conditional on car ownership

Table 5-1: Baseline Fleet Composition

	Year 1			Year 10		
Class	New	Used	All cars in operation	New	Used	All cars in operation
Compact	4.98	44.68	49.66	5.27	49.52	54.79
Lux compact	0.22	4.44	4.66	0.26	2.79	3.05
Midsize	2.63	27.58	30.21	2.82	27.30	30.12
Fullsize	1.32	16.32	17.64	1.49	14.64	16.13
Lux mid/full	0.32	8.30	8.62	0.39	4.67	5.06
Small SUV	1.32	10.65	11.97	1.41	12.99	14.40
Large SUV	1.10	15.93	17.02	1.30	12.92	14.23
Small truck	1.27	10.26	11.54	1.35	12.25	13.60
Large truck	2.17	19.83	22.00	2.42	22.16	24.58
Minivan	1.32	12.74	14.06	1.45	13.62	15.07
Total	16.65	170.73	187.39	18.15	172.87	191.03

Units are millions of privately owned cars in operation.

Table 5-2: Change in Gasoline Use with 25 Cent Tax Increase

Recycling Method	Flat		Income-based		VMT-based	
	Year 1	Year 10	Year 1	Year 10	Year 1	Year 10
Baseline gasoline use per household (gallons)	775.18	828.89	775.18	828.89	775.18	828.89
% change in gasoline use	-5.08%	-4.97%	-5.06%	-5.05%	-4.51%	-4.38%
% change in VMT	-5.01%	-4.83%	-4.98%	-4.92%	-4.43%	-4.21%
% change in VMT per car	-4.62%	-4.36%	-4.56%	-4.38%	-4.01%	-3.70%
% change in cars	-0.41%	-0.48%	-0.44%	-0.57%	-0.44%	-0.54%
% change in overall MPG	0.08%	0.15%	0.08%	0.13%	0.09%	0.17%

Table 5-3: Fleet Size and Composition

	Baseline[1]		25-cent gasoline tax increase[2]					
			Flat recycling		Income-based recycling		VMT-based recycling	
	Year 1	Year 10	Year 1	Year 10	Year 1	Year 10	Year 1	Year 10
Cars in operation:								
All	188.3	191.0	-0.41%	-0.48%	-0.44%	-0.57%	-0.44%	-0.54%
New	16.7	18.2	-1.00%	0.00%	-1.12%	-0.37%	-0.93%	-0.04%
Used	171.6	172.8	-0.35%	-0.53%	-0.37%	-0.59%	-0.39%	-0.59%
Low MPG	75.9	78.9	-0.47%	-0.77%	-0.50%	-0.80%	-0.50%	-0.75%
High MPG	112.4	112.1	-0.37%	-0.28%	-0.40%	-0.40%	-0.39%	-0.38%

[1]Millions of cars.

[2]Percent change relative to the baseline.

Table 5-4: Revenue and Costs from Increased Gasolne Taxes (Results for Year 1)

Revenue recycling	Flat			Income-based			VMT-based		
Tax increase (cents)	10	25	75	10	25	75	10	25	75
Net tax revenue ($billion)	7.43	17.96	48.46	7.43	17.98	48.44	7.52	18.29	49.91
Efficiency cost*									
Total ($billion)	1.23	3.26	11.43	1.26	3.32	11.80	1.12	2.93	10.55
Per dollar of additional revenue	0.165	0.182	0.236	0.170	0.185	0.244	0.149	0.160	0.211
Per avoided gallon of gasoline consumed ($)	0.71	0.77	0.96	0.74	0.79	0.99	0.73	0.78	0.99

*Negative of the weighted sum of equivalent variations of each household.

Table 5-5: Consumption, Mileage, and Car-Ownership Patterns of Household Income Groups*

Income Decile	Gasoline Consumption		Miles Traveled		Avg. Fuel-Economy of Owned Vehicles**	Share of Economy's Light Trucks and SUVs
	Avg. Level (gallons)	Share of Total	Avg. Level (000's)	Share of Total		
1	157.3	0.02	4.03	0.02	25.61	0.02
2	315.7	0.04	7.97	0.04	25.25	0.04
3	473.6	0.06	11.69	0.06	24.68	0.05
4	588.3	0.08	14.33	0.08	24.35	0.08
5	724.0	0.09	17.65	0.09	24.38	0.09
6	823.7	0.11	19.76	0.11	23.98	0.11
7	922.0	0.12	22.35	0.12	24.25	0.13
8	1060.8	0.14	25.46	0.14	24.00	0.15
9	1227.1	0.16	29.55	0.16	24.08	0.17
10	1459.8	0.19	35.28	0.19	24.17	0.17

* Predicted values from simulation model
** VMT-weighted

Table 5-6: Decomposition of Welfare Impacts of 25 Cent Gasoline Tax Increase (Results for Year 1)

	Gasoline price	Transfer	Car prices	Producer profits	EV	EV as a percent of income
Flat Recycling						
Income						
<25	-84.36	157.59	2.65	-3.19	74.90	0.45%
25-50	-196.37	160.23	-0.36	-7.34	-52.09	-0.14%
50-75	-284.10	158.89	-3.16	-12.13	-154.97	-0.25%
>75	-334.46	160.30	-4.76	-19.51	-214.82	-0.22%
All	-176.02	159.05	0.07	-7.37	-29.97	-0.08%
Income-based Recycling						
Income						
<25	-83.90	68.33	2.95	-3.51	-13.86	-0.08%
25-50	-196.41	157.21	-0.33	-8.08	-55.75	-0.15%
50-75	-284.66	259.82	-3.44	-13.36	-56.02	-0.09%
>75	-336.04	417.89	-5.31	-21.49	38.71	0.04%
All	-176.07	157.84	0.12	-8.12	-31.85	-0.08%
VMT-based Recycling						
Income						
<25	-84.26	79.41	2.91	-2.89	-2.03	-0.01%
25-50	-197.38	181.57	-0.31	-6.64	-30.87	-0.08%
50-75	-285.90	261.03	-3.35	-10.98	-53.56	-0.08%
>75	-340.01	307.50	-5.19	-17.66	-71.16	-0.07%
All	-177.08	162.94	0.13	-6.67	-26.04	-0.07%

Welfare effects are expressed in price-adjusted dollars.

Table 5-7: Welfare Impact of 25 Cent Gasoline Tax Increase on Selected Household Groups

	Year 1			Year 10		
Recycling	Flat	Income-based	VMT-based	Flat	Income-based	VMT-based
All	-29.97	-31.85	-26.04	-30.76	-32.32	-26.16
Retired	46.45	8.25	-15.37	54.01	12.82	-12.98
Not retired, no children	-26.48	-33.36	-21.93	-21.74	-28.48	-15.82
Not retired, with children	-88.20	-58.76	-38.10	-100.94	-68.64	-46.77

Welfare effects are expressed in price-adjusted dollars.

Table 5-8: Impacts of Gasoline Taxes under Alternative Parameter Assumptions

	Year 1		Year 10	
	Baseline	25 cent tax increase[1]	Baseline	25 cent tax increase[1]
Central Case				
Gasoline consumption (gallons/household)	775.18	-5.08%	828.89	-4.97%
Aggregate VMT (000's miles/household)	18.80	-5.01%	21.23	-4.83%
Avg. MPG (miles weighted)	24.26	0.080%	25.62	0.152%
Avg. EV (price-adjusted dollars)	-	-30.39	-	-31.03
Faster Fuel-Economy Improvements[2]				
Gasoline consumption	773.66	-5.07%	751.56	-4.48%
Aggregate VMT	18.83	-4.99%	22.25	-4.23%
Avg. MPG	24.34	0.078%	29.60	0.258%
Avg. EV	-	-29.93	-	-25.44
High Scrap Elasticity				
Gasoline consumption	775.18	-5.15%	828.89	-4.99%
Aggregate VMT	18.80	-5.07%	21.23	-4.84%
Avg. MPG	24.26	0.085%	25.62	0.152%
Avg. EV	-	-29.95	-	-30.70

[1]Percent change relative to the baseline under the same parameter assumptions.

[2]Percent increases in fuel economy over 10 years are:
Compact 41, Lux compact 41, Midsize 52, Fullsize 58, Lux mid/full 55,
Small SUV 54, Large SUV 65, Small truck 58, Large truck 59, Minivan 59

Welfare Impacts[1] across Household Income Groups
Under Alternative Revenue-Recycling Methods

Figure 5-1a: Year 1, 25 Cent Tax

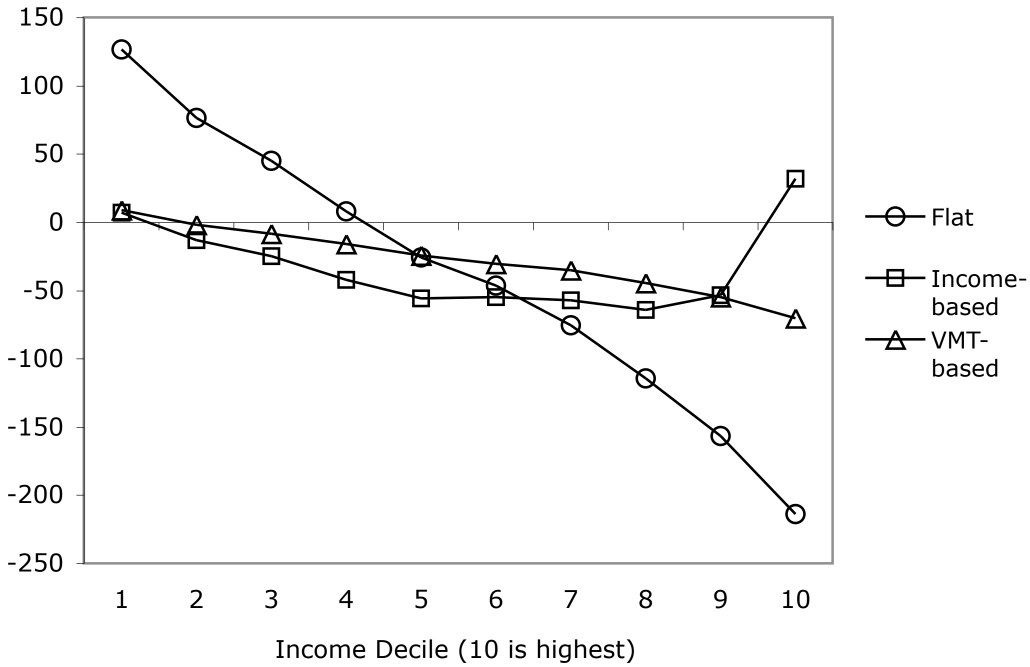
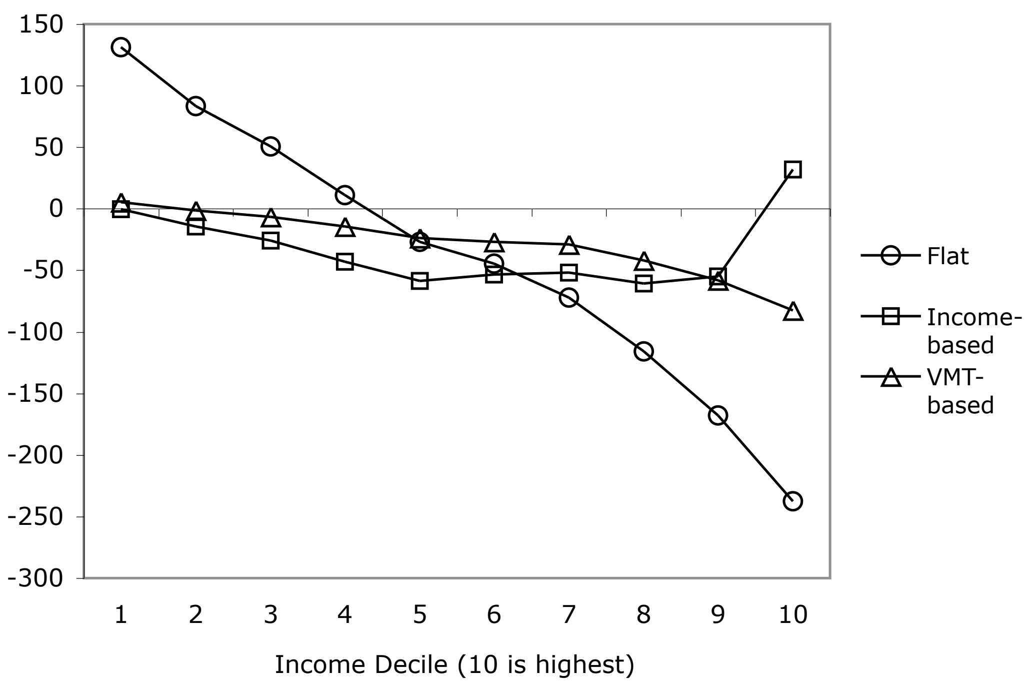


Figure 5-1b: Year 10, 25 Cent Tax



[1]Welfare impacts are in average price-adjusted dollars per household.

Figure 5-2a: Base VMT by Race and Income

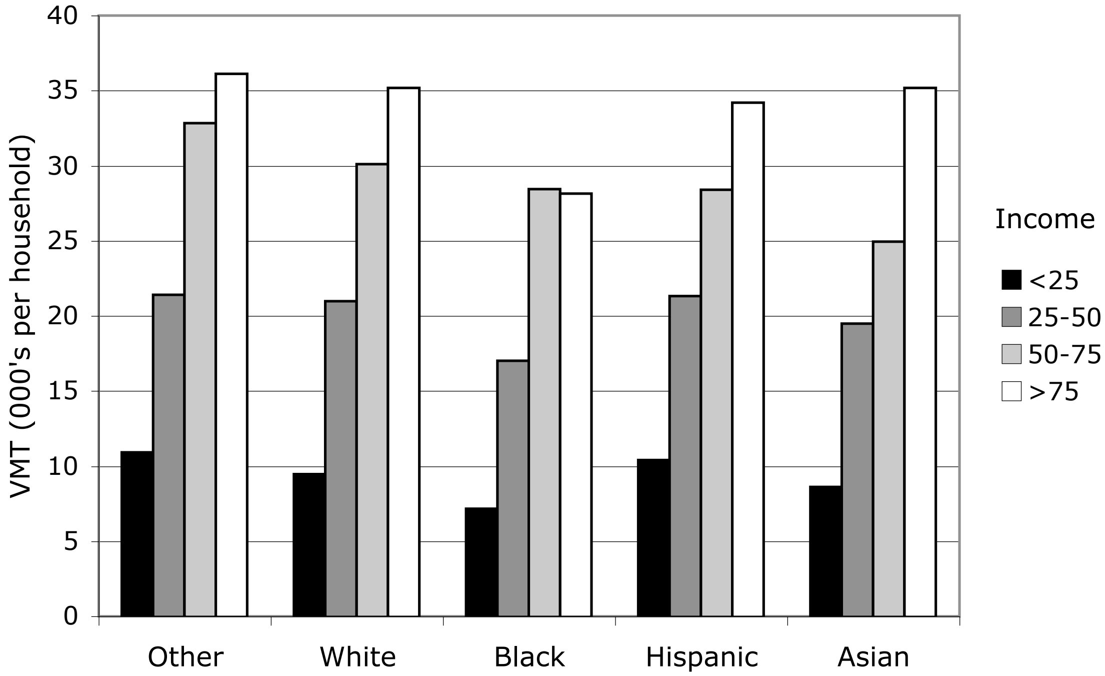


Figure 5-2b: Household EV by Race and Income - 25 Cent Gas Tax Increase with Flat Recycling

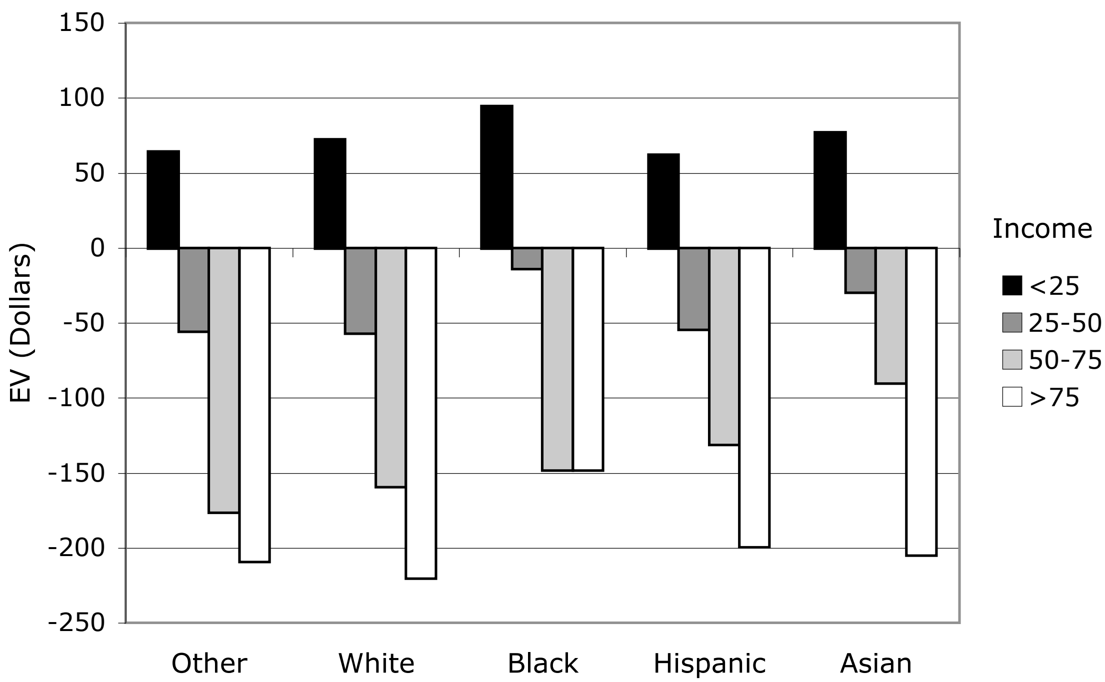


Figure 5-3a: Average Household VMT by State

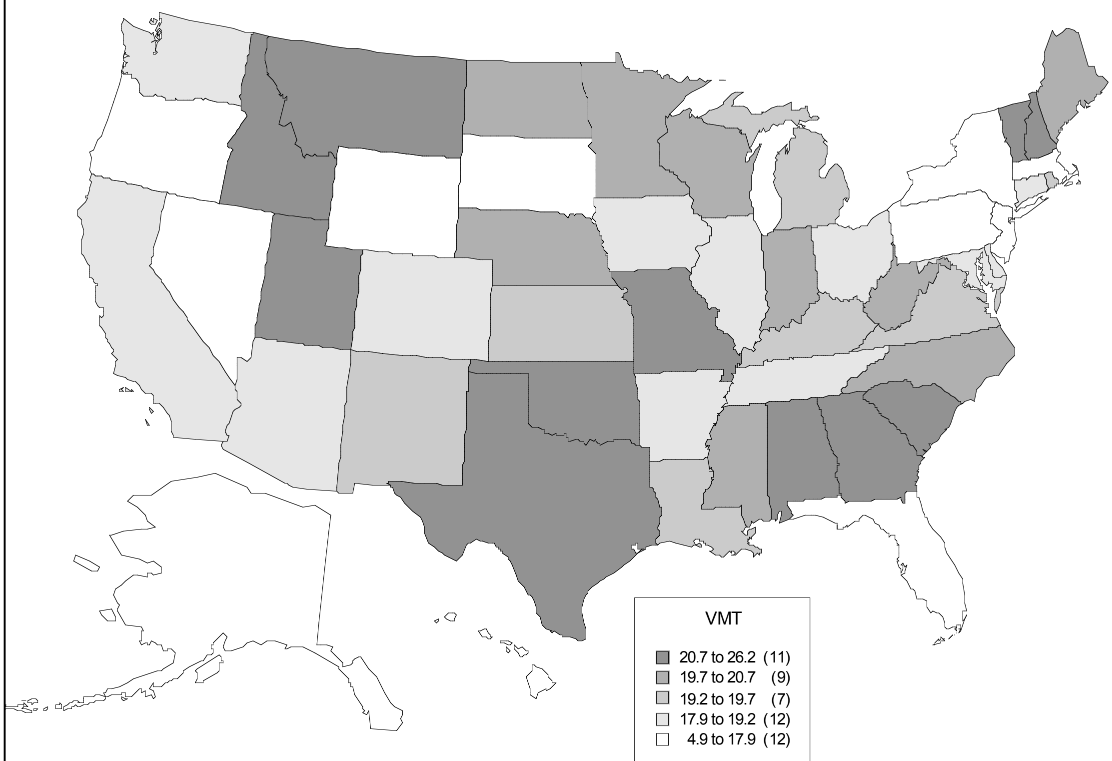


Figure 5-3b: Average Household EV --25 Cents Gasoline Tax With Flat Recycling--