A Functional Form Analysis of the Short-Run Demand for Travel and Gasoline by One-Vehicle Households

DAVID L. GREENE AND PATRICIA S. HU

The short-run elasticity of vehicle travel and gasoline demand is analyzed using gasoline purchase diary data for households in the United States owning one vehicle. A Box-Cox method (iterative ordinary least squares) is used to determine best functional forms for each of four income quartiles in the sample. Transformation parameters for all income groups are found to be close to 0.4. Thus, price elasticity increases with increasing fuel prices. Elasticity estimates at the mean for the three upper quartiles are -0.6, and that for the lowest quartile is approximately -0.5.

How the short-run price elasticity of gasoline demand varies with income determines how severely price increases, whether as a result of shortages or other causes, will affect consumers in different income groups in the short run. This knowledge is important in formulating strategies for possible petroleum shortages, as well as understanding the impacts of fuel taxes. Despite the very large number of econometric studies of gasoline demand that have appeared over the last 15 years (1, 2), very little is known about how price elasticities vary across income groups in the United States. Dahl (3) examined the variability of aggregate gasoline demand elasticities over time and across countries, as did Wheaton (4). Their results showed no great differences in price elasticities across countries with widely differing average incomes. However, these studies did not use individual household data, and it is not possible to extend their results for whole countries to apply them to individual households in the United States. Although household survey data have been used in some studies (5–7), the question of the variability of elasticities across income groups has not been addressed.

The stability of the price elasticity of gasoline demand across income groups is studied using techniques that simultaneously estimate the appropriate functional form. U.S. studies of the functional form of aggregate gasoline demand functions using the Box-Cox method have suggested that the double-log model is appropriate (2, 8). A New Zealand study revealed that results varied depending on whether monthly, quarterly, or annual data were used (9). The analyses are extended in this paper by using disaggregate household data, and equations for income groups are estimated. The data is derived from a gasoline purchase diary survey conducted between April 1978 and March 1981 (10). In order to simplify the analysis, only households owning one vehicle were included in this study. Future work will extend the analysis to multivehicle households.

The derivation of a gasoline demand equation from the household production theory of consumer demand is the subject of the next section; the data used in estimating the demand functions is briefly discussed in the third section; the Box-Cox transformation technique is the subject of the fourth section; estimation results are presented and discussed in the fifth section; and a conclusion follows in the sixth section.
THEORY

Household demand for gasoline is derived primarily from the demand for highway vehicle travel. The economic theory of household production (11, 12) provides an appropriate theory from which to derive models of travel (13, 14), and gasoline demand. Households purchase the necessary inputs for the production of travel (vehicles, gasoline, maintenance, parts) and supply their own labor time to produce the quantity of travel desired. The crucial aspect of household production theory, as opposed to classical demand theory, is that it recognizes the central role of the technology of household production in determining demand. As will be shown below, the key factor in the technology of production that affects gasoline demand is the technical efficiency of the vehicle stock (e.g., miles per gallon). Especially in the short run, when characteristics of the vehicle stock are fixed, fuel economy of the vehicle stock is a critical explanatory variable.

Suppose that the household has the following utility function, $U$, which is weakly separable in vehicle travel, $T$, and a composite good, $z$, then

$$ U = U(z, T) \quad (1) $$

This does not mean to say that the household derives utility directly from travel, but rather that utility is some weakly separable function of travel. The assumption of weak separability is not necessary to the final estimation of demand equations, but is used here to simplify the exposition. The household's economic problem is to maximize $U$, subject to constraints on income and leisure time available for producing travel. Following Michael and Becker (12) these are collapsed into a single constraint on full income:

$$ I = wL - pz - C = 0 \quad (2) $$

where $I$ is monetary income, $w$ is the household's valuation of leisure time $L$, time not spent in producing monetary income, $p$ is the price index of $z$, and $C$ is the cost function for producing travel.

The travel cost function is of particular interest because it embodies the technology for producing travel:

$$ C = [(p_g/\text{mpg}) + v + (w/\text{mph})]T + rA \quad (3) $$

The quantity in parentheses consists of the variable costs of travel: the gasoline cost per mile [price divided by miles per gallon (mpg)], other variable costs, $v$ (e.g., maintenance, lubricants, parts, insurance, etc.), and the time cost [the value of leisure time divided by average speed in miles per hour (mph)]. The final term represents the annualized vehicle cost, and is given by the asset price times a constant that is a function of the household's time discount rate (or its effective interest rate for capital). The technology of energy consumption enters through the determination of the fuel cost per mile.

The household's optimization problem can be written using a LaGrange multiplier:

$$ \text{Max } U^* = U(z, T) - m((I + wL - pz)
- [(p_g/\text{mpg} + v + w/\text{mph}) T + rA]) \quad (4) $$

The first order conditions for optimization are

$$ \frac{\partial U^*}{\partial T} = U_T + m(p_g/\text{mpg} + v + w/\text{mph}) = 0 \quad (5) $$

$$ \frac{\partial U^*}{\partial z} = U_z + mp = 0 \quad (6) $$

$$ \frac{\partial U^*}{\partial z} = (I + wL - pz) $$

$$ + [(p_g/\text{mpg} + v + w/\text{mph}) T + rA] = 0 $$

From these the expected result that the ratio of marginal utilities of travel and the composite good are equal to the ratio of their marginal costs is derived:

$$ \frac{U_T}{U_z} = [(p_g/\text{mpg}) + v + w/\text{mph}]P \quad (6) $$

In order to better illustrate how household demand functions for gasoline can be derived from this problem a particular utility function in Equation 4 can be substituted and the demand function for gasoline solved. This is done for the purpose of illustrating certain properties of all demand functions for vehicle travel. In fact, the functional form of $U$ is unknown. In the estimation section below, the Box-Cox technique will be used to identify from a class of functional forms the one that best fits the data. Suppose that the utility function is additive in the logarithms of its argument [such a form is also weakly separable, as assumed earlier (16)]:

$$ U = \ln(z) + \ln(T) + c \quad (7) $$

The first two equations of the first order conditions now become

$$ \frac{\partial U^*}{\partial T} = (bT) + m(p_g/\text{mpg} + v + w/\text{mph}) \quad (7) $$

$$ \frac{\partial U^*}{\partial z} = (aT) + mp \quad (8) $$

Substituting $m = -(a/p)$ into the first condition gives $0 = bT - (a/p) (p_g/\text{mpg} + v + w/\text{mph})$. Then, solving the third condition (the budget constraint) for $z$, representing all monetary variables indexed to (divided by) the price of $z$ with a prime ($'$), and substituting for $z$ in the equation above gives

$$ T = \frac{b (p_g/\text{mpg} + v' + w'/\text{mph})}{a(p_g/\text{mpg} + v' + w'/\text{mph})} \quad (9) $$

What remains is to solve for the demand for gasoline, instead of the demand for travel. A particularly simple way to represent the household production function for travel is as a fixed input production function. That is, constant proportions of each input must be supplied (17):

$$ T = \min(g_{\text{mpg}}, h_{\text{mph}}, n_{\text{veh}}, \ldots) \quad (9) $$

where the first three inputs represent gasoline, time spent driving, and vehicle stock. At least in a short-run situation, a very good argument can be made that motor fuel is the only significant variable input to vehicle travel. Therefore, short of a breakdown, maintenance can be considered an annual cost and capitalized along with the cost of vehicle ownership. In any case, gasoline is always a limiting factor so that

$$ T = \min(g_{\text{mpg}}, \ldots) = g_{\text{mpg}} \quad (10) $$

is always true in the short run.
function was arbitrarily chosen, yet it illustrates many important features of any valid gasoline demand equation:

\[ g = \frac{1}{1 + \frac{(a/b)}{1/mpg \cdot (T' + w/L - rA')}} \]

First, gasoline demand is a function of household income and the value of household time. As income increases, demand for gasoline will increase \((a, b > 0 \text{ is implicitly assumed})\). As the value of household leisure time, \(w\), increases, it affects gasoline demand in two opposing ways. Directly as the value of time increases, it increases full household income, which tends to increase the demand for gasoline. On the other hand, it also increases the cost of household labor used to produce travel, which tends to decrease the demand for gasoline. Fuel economy has similar opposing effects. If higher fuel economy did not reduce the cost of travel, then an increase in miles per gallon would result in a proportional decrease in the demand for gasoline. Increased miles per gallon, however, also reduces the variable cost of travel, thereby increasing the demand for gasoline, because demand is inversely related to the total variable costs of travel.

The price elasticity derived from Equation 11 is not constant but depends on price, miles per gallon, the time cost of travel, and therefore on the wage rate. Although the particular form of the elasticity is peculiar to this example (Equation 12), it does illustrate that price elasticities are generally not constant.

\[ e_p = - \frac{1 + [(v' + w/\text{mph}) \cdot \text{mpg}]}{1/\text{mpg} - \text{w}/\text{mph}} \]  

This simple demand equation reveals a great deal about household gasoline demand and provides useful guidance about how to structure an equation that can be calibrated using actual consumption data. However, there are several important issues that still remain to be addressed. First, a particularly simple functional form was chosen for purposes of exposition. In estimation, allow for the possibility that other mathematical formulations better fit the data. This is taken up in the fourth section. Second, allow for the possibility that even in the short run the adjustment process may be dynamic. Because monthly data will be used to estimate the model, this is equivalent to saying that full, short-run adjustment occurs within a month. There is some empirical support for this assumption (18), yet it would be an improvement to test for its validity.

Finally, there is the problem of including other relevant household characteristics in the demand equation. Factors such as the number of licensed drivers, the spatial environment in which the household lives (19), and the availability of alternative transportation modes are clearly relevant. Because interest centers on the price elasticity of demand, these aforementioned influences need to be controlled to the greatest extent possible or removed. This is achieved by "centering" each household's data about the household mean and is described in the fourth section.

### DATA

The source of the data used in this study is the National Family Opinion Poll (NFO) Gasoline Diary Panel survey, as modified by the Energy and Environmental Analysis (EEA), Inc. (10). The original Gasoline Diary Panel consists of approximately 734,000 fuel purchase records completed for more than 15,000 vehicles over a 36-month period from April 1978 to March 1981. Each respondent entered data on each fuel purchase for every vehicle owned. The data included fuel type, odometer reading, gallons purchased, cost of purchase, and price paid per gallon. In addition, household demographic and economic data were recorded for each household at the beginning of its participation in the survey.

These individual purchase records were collapsed by EEA into a monthly summary data file about one-fifth the size of the original data base. In cases where purchases straddled 2 months, mileage and fuel consumption were allocated proportionally to time. From this data base all those households owning only one vehicle were selected. This left a total of 3,777 households with 46,256 total monthly observations. This sample was subdivided into rough quartiles: (a) under $8,000/year, (b) $8,000 to $12,000/year, (c) $13,000 to $19,000/year, and (d) over $19,000/year. Each quartile contained roughly 1,000 households and 10,000 monthly observations. In each quartile 5 to 10 percent of the observations were either missing or unusable for some other reason.

Summary data for the income quartiles show a clear relationship between income and vehicle travel, as well as expenditures on fuel. Not only does the highest income quartile travel about 250 mi per month more than the lowest, but they also pay about 10 percent more for the fuel they buy (Table 1). Interestingly, there is very little variation across income groups in average miles per gallon (total miles per total gallons). Fuel prices are deflated to 1967 dollars using the consumer price index for urban consumers.

An important fact about the NFO survey data is that it is a representative panel survey, not a statistically valid random sample. Furthermore, it should be kept in mind that data for single-vehicle households only were used in this study.

### TABLE 1 SUMMARY STATISTICS FOR INCOME QUARTILES

<table>
<thead>
<tr>
<th>Quartile</th>
<th>N</th>
<th>Mean Monthly Travel (mi)</th>
<th>Mean Fuel Price (c/gal)</th>
<th>Mean Fuel Expenditures (c/gal)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest</td>
<td>10,195</td>
<td>706</td>
<td>0.405</td>
<td>17.55</td>
</tr>
<tr>
<td>Low-mid</td>
<td>12,810</td>
<td>828</td>
<td>0.419</td>
<td>17.40</td>
</tr>
<tr>
<td>High-mid</td>
<td>10,466</td>
<td>934</td>
<td>0.420</td>
<td>17.66</td>
</tr>
<tr>
<td>High</td>
<td>9,674</td>
<td>964</td>
<td>0.440</td>
<td>17.56</td>
</tr>
</tbody>
</table>

*For 1967.
ESTIMATION

In order to estimate demand equation parameters, a specific functional form must be chosen. Choice of functional form, however, can affect both the point estimates of price elasticity and how those estimates may vary as a function of price and quantity consumed. As a result it is desirable to use a method that allows functional form and parameter estimates to be inferred from the data. The most generally accepted and widely used technique for inferring functional form is by Box and Cox (20). The Box-Cox approach allows generalized linear models of the form

\[ g(\lambda_i) = b_0 + b_{1x1}(\lambda)x_1 + \ldots + b_{kxk}(\lambda)x_k + e \]  

(13)

where the \( x_i \)'s are explanatory variables, \( e \) is a random error term, and the \( b_i \)'s are parameters to be estimated. The Box-Cox power transformation is defined as

\[ g(\lambda) = \frac{(g - 1)/\lambda}{\lambda \neq 0} \]

\[ \ln(g) \quad \lambda = 0 \]  

(14)

The functional form of the equation is dictated by the \( \lambda_i \)'s, which are estimated at the same time as the \( b_i \)'s. Spitzer (21) describes four equivalent methods for estimating the parameters of a Box-Cox model. The method used here is iterative ordinary least squares (IOLS) using the scaling trick proposed by Zarembka (22). By multiplying Equation 13 through by the geometric mean of observations on the dependent variable raised to the \(-\lambda\), \( g^{-\lambda} \), the \( R^2 \) of the scaled regression can be used to determine the value of \( \lambda \). Actually, the dependent variables need not be scaled by \( g^{-\lambda} \); provided it is recognized that the estimated \( b_i \)'s will be scaled accordingly:

\[ b_i^* = g^{-\lambda}b_i \]  

(15)

The IOLS method is so called because regressions are iteratively performed using different values of \( \lambda_i \)'s until the \( \lambda_i \)'s providing the highest \( R^2 \) are found to the desired degree of accuracy.

In addition to the cost per mile (cpm) of motor fuel, a number of household and vehicle-specific variables might be expected to be included in Equation 13 as explanatory variables. Factors such as income, number of drivers, location of the household, availability of transit, population density, and others are candidates. The survey contains data for some of these but not for others. In addition, from Equation 10 the entire right-hand side of Equation 13 should be divided by miles per gallon.

To avoid many of the complications described, two tricks were used. First, monthly gasoline consumption was multiplied by average miles per gallon for the household and month, so that the dependent variable became monthly travel (in fact, miles per gallon in the survey were calculated as travel divided by fuel consumption, so that this is identical.) Second, after scaling and transformation, the data were centered by subtracting the household mean from each household's observation. This trick allows all household-specific variables that are constant over time—this includes income because monthly income was not recorded—to be dropped from the equation. It also eliminates the intercept term; in effect, each household has its own, unestimated intercept. The only variable remaining on the right-hand side is fuel cost per mile. It is possible that non-household-specific, time-dependent variables should be included in the regression, such as seasonal dummy variables or national economic trend variables. No attempt was made, however, to test such hypotheses in this analysis.

The final estimating form of the regression equation is therefore

\[ g(\lambda) = b^*(\text{price/mpg})^{(\lambda)} \]  

(16)

Although the scaling and centering considerably simplify the regression equation to be estimated, they pose significant data processing burdens in sorting, calculating means, and transforming thousands of data observations. In order to reduce the number of iterations and hold down the cost of the analysis, only a single transformation parameter was estimated, \( \lambda \). The 1982 PROC REG procedure of the SAS Institute was used throughout.

RESULTS

Estimation of the travel demand equations produced remarkably consistent results across income quartiles. Three of the four independently estimated IOLS solutions resulted in values of \( \lambda \) very close to 0.4. Only the lowest income quartile differed. Its transformation parameter was closer to 0.3, but the \( R^2 \) function is very flat in the region between 0.2 and 0.4 (Figure 1). It appears that the functional form of the relationship between vehicle travel and fuel cost per mile, and therefore, fuel use and fuel cost per mile, is very similar for all income groups. The low values of \( R^2 \) seen in Figure 1 are typical of disaggregate data.

Whereas previous studies using aggregate data have found that the logarithmic transformation (\( \lambda = 0 \)) fits the data base best, these results based on disaggregate data indicate that the constant elasticity formulation does not fit the data best. It can be shown that the cost-per-mile elasticity of travel (and gasoline) demand in the simple Box-Cox model is given by:

\[ e = b_1\text{cpm}/[\lambda(b_0 + b_1\text{cpm}^{(\lambda)}) + 1] \]  

(17)

The elasticity of demand therefore depends on the price of fuel, miles per gallon, and the household intercept, which reflects the level of travel by the household. For reasonable values of the coefficients, elasticity will increase as price increases.

To compute elasticities, an intercept term must be computed because none was estimated. Rather than compute intercepts for each household, one intercept for each income group was
calculated using mean values (of the uncentered, transformed variables) for each group, and the relationships:

\[ b_0^* = \hat{T}(\hat{T})^{(\lambda)} - b_1^*(\text{cpm})^{(\lambda)} \]
\[ b_1 = \hat{T}^\lambda b_1^* \]
\[ b_0 = \hat{T}^\lambda b_0^* + \hat{T}^{(\lambda)} \]  

(18)

The data used to make these calculations, together with the estimated model coefficients are given in Table 2. The suffix bar in a variable name indicates an arithmetic mean, while a dot indicates a geometric mean. Costs per mile are in 1967 cents per gallon.

Both the \( b_1 \) coefficients and, to a lesser extent, the estimated elasticities increase in magnitude with increasing income. The pattern of increase is quite interesting. There is a substantial jump in elasticity from the lowest to the low-middle income group, but elasticities remain essentially constant thereafter. Apparently, price responsiveness of gasoline demand is essentially the same for middle and upper income groups, and slightly less elastic for the lowest income group. The computer elasticities range from about \(-0.5\) to \(-0.6\), larger than the range of \(-0.1\) to \(-0.3\) typically found in the literature for short-run gasoline demand (23). In Table 3 elasticities from the double log model are compared with those from the optimal Box-Cox transformation: the differences are small and the pattern is the same. Using the same data set, Greene and Hu (5) estimated a short-run fuel price elasticity of vehicle travel for all income groups combined of \(-0.3\). They included seasonal factors in the regressions. It is likely that the exclusion of such time-dependent factors in this analysis has inflated the cost-per-mile coefficients.

### TABLE 2 TRAVEL ELASTICITY ESTIMATES FOR HOUSEHOLDS BY INCOME GROUP

<table>
<thead>
<tr>
<th>Income Groups</th>
<th>Lower</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Fourth</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>0.2</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
<td>0.4</td>
</tr>
<tr>
<td>TBAR (( \lambda ))</td>
<td>0.0697</td>
<td>0.1379</td>
<td>0.1143</td>
<td>0.1024</td>
<td>0.1092</td>
</tr>
<tr>
<td>CPMBAR (( \lambda ))</td>
<td>0.79</td>
<td>0.89</td>
<td>0.93</td>
<td>0.93</td>
<td>0.97</td>
</tr>
<tr>
<td>CPMBAR</td>
<td>2.29</td>
<td>2.29</td>
<td>2.35</td>
<td>2.34</td>
<td>2.41</td>
</tr>
<tr>
<td>LOG(TDOT)</td>
<td>6.43</td>
<td>6.42</td>
<td>6.61</td>
<td>6.74</td>
<td>6.76</td>
</tr>
<tr>
<td>TDOT</td>
<td>620.17</td>
<td>620.17</td>
<td>742.48</td>
<td>845.56</td>
<td>862.64</td>
</tr>
<tr>
<td>( b_1^* )</td>
<td>-0.419</td>
<td>-0.357</td>
<td>-0.437</td>
<td>-0.431</td>
<td>-0.439</td>
</tr>
<tr>
<td>( b_1 )</td>
<td>-1.52</td>
<td>-4.67</td>
<td>6.15</td>
<td>-6.39</td>
<td>-6.56</td>
</tr>
<tr>
<td>( b_0^* )</td>
<td>0.40</td>
<td>0.46</td>
<td>0.52</td>
<td>0.50</td>
<td>0.54</td>
</tr>
<tr>
<td>( b_0 )</td>
<td>14.54</td>
<td>36.20</td>
<td>40.00</td>
<td>42.01</td>
<td>42.84</td>
</tr>
<tr>
<td>Mean elasticity</td>
<td>-0.49</td>
<td>-0.48</td>
<td>-0.60</td>
<td>-0.59</td>
<td>-0.61</td>
</tr>
<tr>
<td>Predicted annual travel</td>
<td>7,616</td>
<td>8,248</td>
<td>9,599</td>
<td>10,838</td>
<td>11,116</td>
</tr>
</tbody>
</table>
TABLE 3 COMPARISON OF FUEL COST ESTIMATES

<table>
<thead>
<tr>
<th>Income Quartiles</th>
<th>First</th>
<th>Second</th>
<th>Third</th>
<th>Fourth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Double logarithmic</td>
<td>λ = 0.4</td>
<td>-0.49</td>
<td>-0.56</td>
<td>-0.55</td>
</tr>
<tr>
<td>Greene and Hu</td>
<td>-0.48</td>
<td>-0.60*</td>
<td>-0.59</td>
<td>-0.60</td>
</tr>
<tr>
<td></td>
<td>-0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*For all.

CONCLUSIONS

This preliminary analysis of the short-run price responsiveness of household motor fuel demand suggests that demand functions vary little across income groups, either in functional form or price elasticity. For the upper three income quartiles, there is nearly complete agreement. Transformation parameters are nearly identically equal to 0.4, and the estimated elasticities at the mean are all -0.6. Only the lowest income quartile differs, and still the differences are slight. Lambda is approximately 0.3, and the price elasticity is somewhat lower, about -0.5. The differences between the lowest and highest income quartiles suggest that the lowest income quartile should be examined more closely to see whether more extreme differences exist within this quartile. The similarity of results for the upper quartiles suggests that it may be possible to aggregate households in these groups for purposes of studying price elasticities.

It is likely that the price elasticity estimates obtained here are inflated by the failure to include other time-dependent factors such as economic trends and seasonality. Future work should address this issue. The functional forms studied here allow one transformation parameter only. The usefulness of additional transformation parameters should be explored. Finally, this study has used data from households owning one vehicle only. This enabled very simple models to be formulated. Future work must address the majority of households owning two or more vehicles.

ACKNOWLEDGMENT

Research was sponsored by Policy, Safety and Environment, Office of Policy, Planning and Analysis, U.S. Department of Energy, under Contract DE-ACO5–84OR21400 with Martin Marietta Energy Systems, Inc.

REFERENCES