Household Gasoline Demand in the United States

by

Richard Schmalensee and Thomas M. Stoker

MIT-CEEPR 95-007WP August 1995

Household Gasoline Demand in the United States

Richard Schmalensee
Massachusetts Institute of Technology

Thomas M. Stoker*
Massachusetts Institute of Technology

August 1995

Abstract

Continuing rapid growth in U.S. gasoline consumption threatens to exacerbate environmental and congestion problems. We use flexible semiparametric and nonparametric methods to guide analysis of household gasoline consumption in 1988 and 1991. The number of licensed drivers has a strong effect on consumption, and including this variable cuts the estimated income elasticity in half. Slower projected future growth in licensed drivers points to slower growth in gasoline consumption. A parsimonious representation of age, income, lifecycle and location effects is developed and tested. We show how flexible methods also helped reveal fundamental problems with the available price data.

* Sloan School of Management, E52-455, Massachusetts Institute of Technology, Cambridge, MA 02139. This research was supported by the MIT Center for Energy and Environmental Policy Research, the U. S. Department of Energy and the National Science Foundation. The authors wish to acknowledge numerous helpful comments from A. Denny Ellerman, Howard Gruenspecht, Edward D. Porter and the audiences of the CEEPR Spring meetings, the Neyman Seminar at the Department of Statistics, University of California at Berkeley and the Institute for Fiscal Studies in London, England.

October 1995

Corrections to
"Household Gasoline Demand in the United States"
by Richard Schmalensee and Thomas Stoker

In order to clarify the statements made in this paper regarding the RTECS data, the following changes will be made in the next revision of the paper:

Page 17 *Replace the second line following the "Gasoline Price = " formula;*

> We investigated price effects using this variable, but ended up concluding that the RTECS data contained no useful information on those effects.
>
> *with*
>
> We investigated price effects using this variable, but ended up concluding that the 1988 and 1991 RTECS data contained no useful information on those effects.

Page 23 *After the first line of the first whole paragraph;*

> Thus all consumers in each region were *assumed* to face the same prices.
>
> *add the parenthetical note*
>
> (In 1985 and earlier surveys, expenditures were taken from diaries, so this assumption was not imposed.)

1. Introduction

Secular changes in household gasoline consumption have had significant effects on total U.S. demand for energy and, some argue, our national security. In 1991, the average U.S. household spent $1,161 for vehicle fuel, and household vehicles accounted for 31 percent of U.S. petroleum consumption and 13 percent of total U.S. energy consumption (EIA (1993)). Between 1966 and 1991, fuel consumption by cars and light trucks increased by 60 percent, despite a 44 percent increase in average fuel economy (Porter and Rao (1993)).[1] That is, vehicle miles traveled increased by 131 percent over this period, even though total population rose by only 29 percent.

The rapid increase in U.S. gasoline demand in the 1970's and 1980's was not foreseen. Ellerman (1993) has compared forecasts of U.S. gasoline consumption made in the 1974 Project Independence Report with actual experience.[2] As Figure 1 shows, the Project Independence Evaluation System (PIES) forecasts for gasoline demand were low, even though they were based on very optimistic economic growth assumptions. As Figure 1 also shows, using actual GDP and other exogenous variables in the PIES model to generate comparable "forecasts" results in a dramatic underprediction of actual consumption. Ellerman shows that this is not a general pattern; in particular, consumption of electricity was significantly *over*predicted by the PIES model.

If rapid growth in U.S. gasoline consumption continues, the problems of urban pollution and congestion can be expected to intensify, and carbon dioxide emissions, which appear likely to contribute to global climate change, will increase significantly.[3] The consequences of continued rapid growth of gasoline consumption outside the U.S. may be even more serious, of course: with only 4.7 percent of the world's population in 1991, the U.S. accounted for about 25 percent of world oil consumption.[4] One interested in projecting trends in gasoline consumption can turn to an extensive econometric literature on gasoline demand, based largely on aggregate data.[5] These studies tend to find a long-run income elasticity of around

[1] About 20 percent of this total is accounted for by commercial vehicles, which are beyond the scope of this study (Source: EIA (1993, p. 37) and Federal Highway Administration, *Highway Statistics*).

[2] FEA (1994); see Hausman (1975) for a contemporary critique.

[3] On climate change, see Nordhaus (1994).

[4] World Bank (1994), British Petroleum (1993).

[5] For surveys of this literature, see Dahl and Sterner (1991) and Dahl (1993).

unity, suggesting substantial future growth in gasoline consumption in the U.S. and abroad.

This study was motivated by two problems that arise when one considers using existing demand studies to project secular changes in gasoline consumption.[6] First, under even moderately optimistic assumptions about productivity growth, after a few decades a large fraction of U.S. and other OECD households will have per capita incomes well above the range of historical national averages on which most existing income elasticity estimates are based. Will households with incomes of $60,000 really drive about twice as much on average as households with incomes of $30,000? If not, the aggregate income elasticity of demand may fall well below unity over time, and consumption growth may slow relative to GDP growth. The most natural way to learn about the gasoline demand of future high-income consumers is to study the behavior of today's high-income consumers. This requires using household-level data.

Second, over periods of a decade or more, age structures and other demographic characteristics may change substantially, and it is reasonable to expect such changes to affect gasoline demand. Unfortunately, as Dahl (1993) notes, very little work has been done on gasoline demand at the household level, and, to our knowledge, none has taken full account of differences in household composition or other demographic dimensions.[7]

The use of household-level cross-section data, which is necessary to investigate the behavior of wealthy households and the influence of demographics, is consistent with our interest in longer-run determinants of demand. We accordingly employ data from the *Residential Transportation Energy Consumption Survey* (RTECS). These data are discussed in the next section, along with our basic modeling framework. We focus on the 1991 data, the most recent available when this study began, and then validate our basic results with the 1988 data and a small panel of households observed in both survey years.

Section 3 presents our results on the Engel structure of demand. We present a

[6] Additional important problems, beyond the scope of the study, stem from difficulties in forecasting changes in household vehicle characteristics, particularly fuel efficiency, and, especially in developing nations, development of road networks.

[7] Hausman and Newey (forthcoming), which we discuss further below, estimate gasoline demand equations at the household level but do not consider the impacts of demographic variables. Jorgenson, Sleznick and Stoker (1988), among others, include demographic variables in models of the demand for energy and other commodities, but that work does not focus on gasoline demand of analyze the structure of income effects in detail.

fairly simple model of income and demographic effects, the specification of which is guided by semiparametric methods of broad applicability and is then validated by various specification tests. We find no evidence that the income elasticity of demand falls at high incomes, but our analysis of demographic effects helps explain why the Project Independence forecasts were too low *and* why demand equations based on recent aggregate data may well overpredict future growth.

These results are also of methodological interest, as practical applications of recently proposed flexible estimation techniques. We compute local average estimates of the gasoline regression surface, and present them graphically. This gives an extremely clear depiction of the income and age structure of gasoline demand in the data; a depiction that is further amenable to making accurate elasticity estimates through parsimonious parametric modeling. In other words, our application provides some guidelines for using nonparametric methods to determine basic structure, and guide subsequent modeling; a practice that we feel will be valuable in many application areas.

Section 4 considers estimation of price elasticities using the RTECS data. Our initial regression estimates were quite plausible, but we ultimately concluded that measurement problems make them uninformative. We discuss this aspect of our study in some detail both to help future investigators avoid misusing the RTECS data and to illustrate the diagnostic value of nonparametric methods, which led us to continue questioning our basic estimates until we ultimately uncovered fundamental measurement problems.

We conclude in Section 5 with a summary of our results and their implications and a brief discussion of the use of vehicle ownership data in models of gasoline demand.

2. The Basic Framework

2.1. The RTECS Data

The *Residential Transportation and Energy Consumption Surveys* (RTECS) are a series of detailed household surveys on driving behavior and vehicle ownership collected by the Department of Energy, beginning in 1979 and now carried out every three years. We focus on the two most recent surveys available when this study began (1988 and 1991), because they are based on the same survey design and data collection methods.[8] We limit our attention to households with non-zero numbers of miles driven, drivers, and cars owned.[9] The resulting data sets are comprised of 2684 household observations in 1991 and 2594 household observations is 1988. By matching household identification numbers, we discovered that 547 households were surveyed in both years, and we report estimates using this short panel below.

Observations on mileage driven in each year were collected directly from odometer readings. These observations were combined with estimated miles-per-gallon figures for each vehicle owned to construct total gallons of gasoline.[10] (There are also observations on annual gasoline expenditures, but we defer discussion of them to Section 4 below).

Households report their annual income in one of 23 ranges (in thousands of dollars). We used a log-normal procedure to estimate within-cell means for each of the ranges, and set each household's observed income to the mean of the appropriate range.[11] We regard the resulting log income values as sufficiently continuous to use them for nonparametric estimation below.

[8]See EIA (1993) for a summary of the statistics from the 1991 survey, as well as a detailed discussion of changes in collection methods and data definitions between our (1988 and 1991) surveys and the earlier ones (1979-1981, 1983 and 1985). The 1988 survey is summarized in EIA (1990).

[9]It was not strictly necessary to drop households with zero vehicles owned, but this eliminated only around 100 observations in each survey, or around 3% of the original samples.

[10]Our interest is in gasoline consumption, but we could have based our estimates on mileage directly, since one can argue that mileage is a better measure of transportation services than gallons. Of course, one can argue the reverse, and, as we discuss below, the results we obtained using mileage driven as the dependent variable are qualitatively identical to the results obtained using gasoline.

[11]Specifically, for each year we estimated the mean and standard deviation of log-income in the population, using the maximum-likelihood method for grouped lognormal data; see Aitcheson and Brown (1963, p.51-2). (The first-order equations were solved iteratively; initial estimates were based on linear approximations to the median and inter-quartile range. Permitting a non-zero bound on household income (and thus a three-parameter lognormal distribution) did not materially improve fit). Using these estimated parameters, we then set each household's income equal to mean income conditional on its reported income range. Experiments with grouping estimators indicated that the resulting measurement error was of negligible importance.

Since the highest income range ($75,000 and above) is unbounded, some distributional assumption of this sort is necessary for these data. In experiments with polynomial models, we employed estimates of higher order conditional moments of log income, but found no substantive differences. This is consistent with the fact that our estimated relationships are quite linear (as discussed below), so that we do not discuss these estimates here.

Location effects — climate and typical modes of driving — are captured via observations on urban, suburban and rural residence as well as regional location. The standard nine region classification of the United States is utilized. Table 2.1 summarizes the main data definitions, where "Base Category" refers to the discrete variables that are omitted from the regression analysis that follows.

Demographic aspects of households are observed in the following ways. Age-of-head is observed in years, which we regard as continuous for nonparametric estimation. Household size (number of household members) and numbers of licensed drivers are observed directly, taking on discrete values (i.e. 1, 2, 3...). Specific ages of each household member are not reported. Rather, each household was assigned to one of nine "lifecycle" categories, described as the "LIFE" variables in Table 2.1. In this way, the survey design attempted to identify different household types associated with different driving needs. Basic summary statistics for the data samples are given in Table A.1.

2.2. The Modeling Framework

We do not employ a tightly parameterized model based on household utility, but the basic framework of our study dictates various aspects of model specification. In particular, we assume that household demand for gasoline arises from an optimization process that involves the number and types of vehicles owned, and the amount of driving for commuting, errands, vacations and personal pleasure. We take as exogenous the basic demographic composition of the household, their (idiosyncratic) tastes for driving and vehicle attributes, and the location in which they live. We also take the wealth of the family as exogenous,[12] and assume that current income is a sufficient statistic for wealth as it affects gasoline demand. Finally, we assume that automobile cost parameters are exogenous, as are available prices for gasoline. For our main work on the Engel structure of gasoline demand, we suppose that price differences are adequately captured in the effects of regional variables.

Given these exogenous features, we assume that each household has chosen the array of vehicles they own — their driving "technology" — as well as (anticipated and realized) amounts of different types of driving. These choices in turn

[12] It is arguable that in certain settings, income will be endogenous, because one may choose a longer commute to a job with a higher salary. We don't take this into account, but it would be a greater concern in a study of gasoline demand in a developing country, where the amount of driving is affected directly by the availability of highways and other infrastructure.

Variable	Description
LTGALS	Log Total Gallons Consumed
LY	Log Income (in Thousands of Dollars)
LDRVRS	Log Number of Drivers
LSIZE	Log Household Size
LAGE	Log Age of Head
RESIDENCE VARIABLES	
URBAN	Urban Residence
SUBURBAN	Suburban Residence (Base Category)
RURAL	Rural Residence
REGION VARIABLES	
REG1	New England (Base Category)
REG2	Middle Atlantic
REG3	East North Central
REG4	West North Central
REG5	South Atlantic
REG6	East South Central
REG7	West South Central
REG8	Mountain
REG9	Pacific
LIFECYCLE VARIABLES	
LIFE1	With Children, Oldest < 7 (Base Category)
LIFE2	With Children, Oldest 7-15
LIFE3	With Children, Oldest 16-17
LIFE4	Two Adults, Age of Head < 35
LIFE5	Two Adults, Age of Head 35-59
LIFE6	Two Adults, Age of Head 60 +
LIFE7	One Adult, Age of Head < 35
LIFE8	One Adult, Age of Head 35-59
LIFE9	One Adult, Age of Head 60 +

Table 2.1: Definitions of Variables

Inputs (Exogenous)	→	Decisions	→	Outputs (Endogenous)
Income		Miles		Total Gallons
Demographics		-Pleasure		Total Miles
Location		- Vacation		MPG
Tastes		-Commuting		
		-Other Driving Needs		
Gasoline Prices				# Cars
		Driving		Car Types
Automobile Cost Parameters		"Technology"		

Table 2.2: Modeling Framework

determine the observed features of driving behavior, namely number of gallons of gasoline, number of miles, and MPG (miles per gallon), as well as the number of cars and car types. In addition to gasoline consumption, this framework indicates how the number of cars owned, types of cars, MPG, etc., are endogenous to the decision process. In this spirit, our main statistical analysis fits a reduced form gasoline demand equation. The main alternative would be to build a complex structural model of vehicle choice and gasoline usage.[13] In this study, we opt for the simplicity of a clear statistical description of the reduced form gasoline demand relationship. Our basic framework is summarized in Table 2.2.

3. The Engel Structure of Gasoline Demand

3.1. Estimates From The 1991 Sample

As described above, the basic household demand model takes the following form:

$$\text{Log Gallons} = F\,(\text{Income, Demographics, Location})$$

We have no predispositions as to the appropriate functional form to use for estimation. While it would be convenient if flexible nonparametric methods could be used to dictate the entire model, this is infeasible given the that there are

[13]A detailed vehicle choice model of this kind is discussed in Train (1986).

22 plausible candidates for predictor variables in Table 2.1. Consequently, some modeling restrictions must be imposed at the outset.[14]

Of our 22 candidates for predictors, two are continuous (Income and Age-of-Head) and the remainder are discrete. Since the U.S. population is expected to get both older and richer over time, we decided to focus flexible modeling methods on the income-age structure of demand. In particular, we begin with a partial linear model that takes log-total gallons as a general function of income and age plus a linear function in the remaining discrete variables. We include income and age in log-form, and likewise for the number of licensed drivers and household size. This semiparametric model is summarized as[15]

$$\begin{aligned} LTGALS &= G\left(LY, LAGE\right) + \beta_1 LDRVRS + \beta_2 LSIZE \\ &\quad + \beta_3' \text{Residence.} + \beta_4' \text{Region} + \beta_5' \text{Lifecycle} + \epsilon \end{aligned} \tag{3.1}$$

The function $G(\cdot)$ has the standard regression interpretation, namely as the income-age structure of mean log-gasoline demand, holding location and other demographic variables constant. We estimate G nonparametrically, and display it graphically.[16]

To discuss estimation, it is useful to rewrite the model compactly as

$$y = G(x) + \beta' z + \epsilon \tag{3.2}$$

where y is log gallons, x denotes log income and log age, and z denotes the remaining variables, and where $E\left(\epsilon|x, z\right) = 0$. Estimation of the coefficients β is based on regressions of (within) variations of y and z around x. Note how

$$E\left(y|x\right) = G(x) + \beta' E\left(z|x\right), \tag{3.3}$$

so that differencing (3.2) and (3.3) yields a linear regression for within deviations:

$$y - E\left(y|x\right) = \beta' \left[z - E\left(z|x\right)\right] + \epsilon. \tag{3.4}$$

[14]While the list includes 18 qualitative variables, a full nonparametric treatment would require accomodation of all possible interactions among those variables as well as with the remaining (continuous and discrete) variables.

[15]Partial linear models were first used in econometrics by Engel, Granger, Rice and Weiss (1986). The estimation method we describe follows Robinson (1988), who demonstrates that the coefficient estimates are $\sqrt{N}$ asymptotic normal, where N is sample size.

[16]Clearly the inclusion of income and age in log-form is inconsequential: the precise income-age structure is estimated regardless of the transformation used for these variables.

In order to implement (3.4) to estimate the coefficients β, we need to use estimates of the regressions $E(y|x)$and $E(z|x)$; namely the mean of y and each component for z for different x values. While virtually any nonparametric regression estimator would do, we use standard kernel estimators, where the estimate of $E(y|x)$ is denoted $\hat{m}_y(x)$ and that of $E(z|x)$ is denoted $\hat{m}_z(x)$.[17] Our estimate of β is just the vector of OLS coefficients $\hat{\beta}$ of the estimated deviation $y_i - \hat{m}_y(x_i)$ on $z_i - \hat{m}_z(x_i)$.[18] Recall that our primary interest is in the income-age structure of gasoline demand, or $G(x)$ in (3.1). Since $G(x) = E\left(y - \beta' z|x\right)$, we could compute $\hat{G}(x)$ as the kernel regression of $y_i - \hat{\beta}' z_i$ on x; the same estimate is given from our earlier kernel regressions as

$$\hat{G}(x) \equiv \hat{m}_y(x) - \hat{\beta}' \hat{m}_z(x) \tag{3.5}$$

The components of the estimate $\hat{\beta}$ are given in Tables 3.1 and A.2.

Let us turn first to a description of the estimate $\hat{G}(x)$. The function $\hat{G}(x)$ can be graphed as a surface over log-income — log-age coordinates. Figures 2a-b present the estimated surface, as well as the estimate of the joint density of log-income and log-age. Some features are evident from these pictures, such as how the lower income households tend to be at the younger and the older ends of the age spectrum. However, not much is clear from the plot of the surface $\hat{G}(x)$. Therefore, we present $\hat{G}(x)$ by drawing cross sections - namely log-income profiles for different ages and log-age profiles for different incomes - as presented in Figures 3a-b.

[17] Since x has two components, the kernel regression estimator is $\hat{m}_y(x) = \hat{f}(x)^{-1} \sum_i \omega_i(x) \cdot y_i$, where $\hat{f}(x) = \sum_i \omega_i(x)$ is the standard kernel density estimator, with $\omega_i(x) \equiv N^{-1}h^{-2}\mathcal{K}\left[\frac{x-x_i}{h}\right]$; c.f. Silverman (1986) and Härdle (1991). For estimation, we standardize the x data, and use the standard normal density for the kernel $\mathcal{K}$. We set the bandwidth as $h = .3$. For reference, given our sample size, the approximate optimal bandwidth for estimating density when x is normally distributed is $h = .258$ (Silverman (1986)), and if the true model is linear, the optimal bandwidth for regression is $h = .294$ (Stoker (1995). Cross validation applied to the residuals from the partially linear model gave $h = .35$, so our estimates may be a bit undersmoothed. However, variation of the bandwidth within these ranges made no difference whatsoever to our estimates (3.5) of the structure of $G(x)$.

[18] Following Robinson (1988), this regression is performed on a trimmed sample, where we omit the 5% of the sample with lowest estimated (x) density. Robinson further notes how the variance of $\hat{\beta}$ is estimated using the usual formulae, without requiring adjustments for the use of the estimates $[\hat{m}_y(x), \hat{m}_z(x)]$ in place of $[E(y|x), E(z|x)]$.

The income structure is quite clear from Figure 3a. First, the relationship appears very linear; except for low income levels, where there appears to be little or no income effect. Second, different income profiles for different ages are roughly parallel, suggesting that the function $G(x)$ is additive in functions of log-income and log-age.

Likewise, the age structure of household gasoline demand is clear from Figure 3b. In particular there is no age effect until age 50, after which gasoline consumption declines smoothly but rapidly. Also, the log-age profiles are also roughly parallel, again reinforcing an additive structure of the basic function $G(x)$.

While these figures depicting $\hat{G}(x)$ tell a clear qualitative story, it is more useful to have a parsimonious quantitative description of the income-age structure, say in terms of elasticities over different ranges. Figures 3a-b suggest using a piecewise linear function in log-income, with different elasticities above and below \$12,000 in household income. For age, we also use a piecewise linear function in log-age, permitting different elasticities above and below age 50. There are several equivalent ways to model these functions: in order to estimate the elasticities in the different ranges, we include the linear spline terms

$$\begin{array}{ll} LY12- & = (LY - \ln(12)) \cdot 1[LY < \ln(12)] \\ LY12+ & = (LY - \ln(12)) \cdot 1[LY \geq \ln(12)] \\ LAGE50- & = (LAGE - \ln(50)) \cdot 1[LAGE < \ln(50)] \\ LAGE50+ & = (LAGE - \ln(50)) \cdot 1[LAGE \geq \ln(50)] \end{array}$$

in place of log-income LY and log-age $LAGE$.[19] The results of these estimations are presented in Tables 3.1 and A.2. We checked whether this parameterization is consistent with the nonparametric estimates using the test statistic of Aït Sahalia, Bickel and Stoker (1994); the value was 1.54, with a (standard normal) p-value of .061. Consequently, we fail to reject the spline specification against the semiparametric partial linear model.[20]

[19] With a constant in the equation, it would be equivalent to include LY and $LAGE$ with one each of the spline terms above, in which case the coefficients on the spline terms would measure the differences in elasticities across the ranges.

[20] This test is a goodness-of-fit test based on $N^{-1}\sum_i I_i [\hat{g}(x_i) - \tilde{g}(x_i)]^2$ corrected for nonparametric bias, where I_i indicates trimming of the 5% of sample values with lowest estimated density. $\tilde{g}(x_i)$ is the fitted value of the final spline regression computed on the trimmed sample. This is a very sensitive test and we view the failure to reject as strong confirmation of the spline specification.

LTGALS - Dependent Variable	Semiparametric Model	OLS Estimates	
LY12-		.024	
		(.043)	
LY12+		.200	.204
	$\hat{G}(LY, LAGE)$	(.020)	(.019)
LAGE50-		.013	
		(.077)	
LAGE50+		-1.32	-1.31
		(.172)	(.164)
LDRVRS	.595	.601	.602
	(042)	(.042)	(.042)
LSIZE	.152	.128	.127
	(053)	(.053)	(.053)
Residence	X	X	X
Region	X	X	X
Lifecycle	X	X	X

Table 3.1: Principal Coefficient Estimates: 1991

Note: "X" signifies inclusion of the relevant set of qualitative variables.

In this fashion, we have used the semiparametric estimates to guide the specification of the parametric model.[21] We also did some OLS exploratory analysis with other income and age terms and interactions, but could find no significant effects.[22] Finally, we carried out the same kind of analysis with the two components of log-gallons, namely log-miles and log-gallons-per-mile. Log-miles exhibits exactly the same qualitative structure as log-gallons, both in the semiparametric specification and the parametric model. Log-gallons-per-mile displays a slight increase with income and age, but the increase is very minor relative to the changes in log-miles.[23] Thus, our results on gasoline demand primarily reflect systematic differences in driving patterns, rather than in vehicle characteristics.

We performed standard F-tests of refinements to the structure of the residence, region and lifecycle effects, and discovered a very parsimonious representation. In all specifications, urban households drive less than suburban households, who in turn drive less than rural households. Households drive less than the norm only in the Northeast and Pacific regions, and the only significant lifecycle effect was that young single adults drive more. These refinements did not involve significant differences from the general linear regression, in view of the size of the sample (an F statistic of 1.69 with a p-value of .046 from an F(15, 2641) distribution), however they were nominally rejected against the semiparametric model with the Aït-Sahalia, Bickel, Stoker (1994) test. The latter test statistic is quite sensitive, and we are not aware of comparative work to assess the practical importance of this latter testing result.[24] In any case, we present these results in Table 3.2. The

[21] We are open to the criticism of pretesting here, of course, but our aim is more to summarize the data than to test a priori parametric hypotheses. Moreover, we are unaware of work that deals with choosing specifications from looking at pictures, as we have done.

[22] Further exploration with very general specifications did lead to one nominal rejection. For the basic model, we allowed all coefficients to vary by the nine lifecycle categories, and the test of no differences gave an F-statistic of 1.43. This is associated with a p-value of .002 from an F(118, 2515) distribution. The resulting model has 143 parameters, with very few of them estimated precisely, and we could not find a useful summary what lifecycle differences were predominant. In any case, we are not greatly concerned with this finding, which does not incorporate much penalty for a huge number of parameters.

[23] We have omitted these results out of a concern for brevity, but they are available on request from the authors.

[24] The main difference concerned a lower estimate of the decrease in consumption for households with age-of-head beyond 50. The nonparametric test may be sensitive to the fact that the refined coefficient conflicts with the estimated age profile with full additive effects. Again, we did not isolate the source of the testing differences.

LTGALS Dependent Var.	1991	s.e.	t-stat.
LY12+	.211	(.018)	(11.4)
LAGE50+	-.913	(.081)	(-11.2)
LDRVRS	.620	(.037)	(16.7)
LSIZE	.097	(.028)	(3.42)
URBAN	-.172	(.026)	(-6.73)
RURAL	.109	(.027)	(4.09)
NON NE ($\sum$REG3-9)	.151	(.027)	(5.52)
REG9 (Pacific)	–.089	(.030)	(-2.93)
LIFE7 (Single Adult, <35)	.183	(.056)	(3.24)
Constant	6.127	(.041)	(149.)
R^2	.396		
N	2684		

Table 3.2: Refined Engel Structure Estimates: 1991

simplicity of this final specification was greatly aided by the accuracy with which we were able to characterize the age-income structure.

The coefficient on log-drivers is extremely robust; it was around 0.6 for essentially all specifications that included log-size and was always precisely estimated. We can always strongly reject the hypothesis of no effect (a zero coefficient) as well as the hypothesis that the appropriate model would be based on consumption per driver (a coefficient of unity). Similarly, the coefficient of log-size is generally around 0.1 and is precisely estimated.

These results have important implications for the history and future of U.S. gasoline demand. Over the 1966-1991 period, the number of licensed drivers increased roughly twice as fast as the population as a whole, both because the driving-age population increased as a fraction of the total population and because an increasing fraction of the driving-age population were licensed.[25] While it seems obvious that this demographic shift would affect the demand for gasoline, it has been ignored in virtually all previous studies of gasoline demand.[26] Our estimates suggest that this shift played a major role in increasing gasoline consumption over the last few decades and that income changes played a much smaller role than most studies suggest. Indeed, in OLS estimation, adding log-drivers generally cuts the estimated coefficient of log-income roughly in half. Since this demographic shift has essentially been completed in the U.S., we can expect future growth in aggregate income and population to produce smaller increases in gasoline demand than in the past.[27] This conclusion is, of course, reinforced by our finding that gasoline demand is lower, all else equal, for household with older heads.

3.2. Estimates from the 1988 Sample and the Household Panel

We carried out the same analysis with the 1988 sample and found a virtually identical income and age structure for gasoline demand. The results for the basic models are presented in Tables A.2 and A.3. The only qualitative differences in the estimates of the basic model were a somewhat smaller estimated effect of household size, and a slightly smaller estimate effect of the decrease in consumption for households with age-of-head greater than 50 These differences are minor, as indicated by the failure to reject the hypothesis that the basic model coefficients are the same in 1988 and 1991 — the F-statistic is .860, with a p-value of .655 from an $F(23, 5258)$ distribution. We carried out the same sort of refining procedures as for the 1991 model, and found slightly different regional structure. The estimates for the refined model are given in Table 3.3, and the F-statistic of the coefficient restrictions used in the refined model is 1.63, which has a p-value of .060 for an $F(15, 2569)$ distribution. In any case, as a matter of validation of the

[25] Both of us had grandmothers who never learned to drive, but our children can't say the same.

[26] After this study was essentially complete, we learned of two exceptions: Gate (1990) and Porter and Rao (1993), both of which employ aggregate data. Porter and Rao (1993) work with vehicle miles per licensed driver. Gately (1990) explains annual aggregate U.S. vehicle miles over the period 1966-1988 with an equation in which the number of licensed drivers appears on the right. He obtains an elasticity of mileage with respect to drivers of 0.65 and finds, as we do, that including the number of drivers cuts the estimated income elasticity roughly in half.

[27] Both Gately (1990) and Porter and Rao (1993) incorporate this effect in medium-term forecasts of gasoline demand.

LTGALS Dependent Var.	1988	s.e.	t-stat.
LY12+	.180	(.019)	(9.55)
LAGE50+	-.901	(.085)	(-10.6)
LDRVRS	.657	(.038)	(17.4)
LSIZE	.084	(.030)	(2.77)
URBAN	-.117	(.025)	(-4.67)
RURAL	.099	(.027)	(3.67)
NON NE ($\sum$REG3-9)	.121	(.027)	(4.49)
REG7 (WSC)	.115	(.035)	(3.30)
LIFE7 (Single Adult, <35)	.122	(.055)	(2.18)
Constant	6.14	(.041)	(150.)
R^2	.387		
N	2594		

Table 3.3: Refined Engel Structure Estimates: 1988

basic model, we are quite encouraged by the fact that the results are so similar between 1988 and 1991.

As mentioned before, a small panel of 547 households was observed in both years. For a final test of model validity, we estimated the refined model in differenced form using this sample, omitting the location variables because none of these households changed locations. The results are presented in Table 3.4.

Qualitatively, the income-age structure is similar, with a smaller estimated income elasticity. Also, the elasticity for number of drivers is smaller, as is the elasticity for number of household members. Since driving habits may react with a lag to changes in household composition, the differences between the panel and cross-section estimates seem quite plausible

For testing, instead of making an assumption about the correlation structure of the disturbances in both years, we tested whether the coefficient estimates of the differenced model were equal to the (estimated) values given in Table 3.2 for the refined 1991 model. This test resulted in an F-statistic of 2.13, with a p-value of .039 from an $F(7,539)$ distribution. The source of greatest difference in fit concerned the log-drivers and log-size effect — for instance, testing all restrictions except for the restriction on log-size gave an F-statistic of .164, with a p-value of .133 for an $F(6, 539)$ distribution. In any case, we feel that the panel estimates are reasonably consistent with our earlier findings, and find no reason to doubt the basic specification of our gasoline demand equations.

4. Estimation of Price Effects

The omission of price effects in the above analysis is unlikely to bias coefficients of included variables because we have controlled for residential and regional differences that are likely to capture much of the variation in retail gasoline prices. Each of the RTECS surveys reports values of total expenditure on gasoline, and one may use these values to try to estimate price effects. In particular, we construct an observed price for each household by

$$\text{Gasoline Price} = \frac{\text{Total Expenditure}}{\text{Total Gallons}}.$$

We denote the log of this price value as LP in the following. We investigated price effects using this variable, but ended up concluding that the RTECS data contained no useful information on those effects. Here we present our analysis because it involved interesting use of nonparametric methods, although our final conclusions are negative.

Part of the motivation for our interest in studying the price effects in this data arise from a recent study of household gasoline demand by Hausman and Newey (forthcoming). This study estimated consumer surplus from tax changes using a data set of roughly 18,000 observations that was constructed by pooling several (1979-1988) earlier RTECS samples. Their model gave nonparametric treatment to price and income structure, and included year and region effects, but no household demographic variables. We include as Figure 4 their estimates of the demand curve at the mean income level, computed using kernel regression and B-spline regression with 6-8 knots.[28]

[28]We thank Whitney Newey for providing us with these figures. Estimators based on B-spline

DLTGALS - Dependent Variable	Differenced Model	
DLY12-	.167	
	(.107)	
DLY12+	.112	.129
	(.049)	(.047)
DLAGE50-	.004	
	(.139)	
DLAGE50+	-.989	-1.02
	(.340)	(.307)
DLDRVRS	.445	.460
	(.090)	(.089)
DLSIZE	.049	.043
	(.083)	(.081)
DLIFE7	.157	.184
	(.131)	(.129)
Constant	-.051	-.053
	(.026)	(.026)
R^2	.135	.131
N	547	547

Table 3.4: Estimates With Differenced Data From Household Panel

Our interest was aroused by the shapes of these curves. In particular, there are increases in quantity demanded over certain price ranges, with no effect (or decreases) over other price ranges. While this is a nonparametric depiction of the demand relation, it seems unlikely that an individual household would be sensitive to price changes in certain ranges but not sensitive to price changes in the center of the overall range. These patterns seem to us to be reflecting some sort of heterogeneity that is not accounted for in the Hausman-Newey model. We began our analysis to see whether our more recent data exhibited such differential reactions, and if so, whether household demographic composition could account for them.[29]

We begin by looking at some OLS estimates of price elasticity. Table 4.1 gives the results of including log-price in log gasoline regressions, both without and with demographic variables. These estimates are all in a range thought to be typical for gasoline elasticities, namely -.8 — -1.1; except for when regional effects are included, when the results become erratic.[30]

While these OLS estimates appear reasonable, a closer look reveals some serious problems. Table 4.2 contains estimates of average derivatives of log-income and log-prices, which are nonparametric estimates of the average of income and price elasticities over the samples. In particular, ADE refers to the average derivatives of log-income and log-price

$$\delta_{LY} \equiv E\left[\frac{\partial G^*}{\partial LY}\right]; \quad \delta_{LP} \equiv E\left[\frac{\partial G^*}{\partial LP}\right]$$

where demographic variables are included in a partial linear way:[31]

$$\begin{aligned} \text{Log Gallons} &= G^*(LY, LP) + \beta_0 \cdot LAGE50 + \ +\beta_1 \text{Log Drivers} \\ &\quad +\beta_2 \text{Log Size} + \beta_3' \text{Lifecycle} \ + \beta_4' \text{Residence} + \epsilon \end{aligned} \tag{4.1}$$

ADE(dw) refers to "density weighted" average derivatives, or the weighted average

approximations are discussed in Chui (1988).

[29]If different demands for different types of households have been mixed in the estimation, it may not be a problem for the consumer surplus calculations. In particular, the consumer surplus estimates might measure average welfare change over the different household types.

[30]For this reason, most of our estimates below omit regional effects.

[31]We fit a partial linear model with three arguments; LY, LP and $LAGE$, and found similar age patterns as before — we didn't find any potential bias problems from including age effects via the age spline term in the additive part of the model.

LTGALS Dep. Var.	1991				1988			
LP	-1.13 (.200)	-.801 (.178)	-.720 (.172)	-.291 (.196)	-1.01 (.189)	-.791 (.167)	-.805 (.162)	.321 (.342)
LY	.330 (.015)	.179 (.015)	.135 (.015)	.160 (.015)	.323 (.016)	.159 (.015)	.121 (.015)	.151 (.034)
LDRVRS		.628 (.038)	.658 (.042)	.612 (.041)		.629 (.038)	.680 (.045)	.049 (.099)
LSIZE		.166 (.027)	.095 (.053)	.123 (.053)		.161 (.029)	.030 (.054)	.293 (.119)
Log-Age			X	X			X	X
Lifecycle			X	X			X	X
Residence			X	X			X	X
Region				X				X

Table 4.1: Various OLS Price Elasticity Estimates

LTGALS Dep. Var.	1991			1988		
	OLS	ADE	ADE (dw)	OLS	ADE	ADE (dw)
LP	-.720	.179	.402	-.805	-1.80	-2.11
	(.172)	(.399)	(.522)	(.162)	(.394)	(.686)
LY	.135	.165	.166	.121	.165	.165
	(.015)	(.018)	(.022)	(.015)	(.018)	(.025)
Demographic	X	X	X	X	X	X
Residence	X	X	X	X	X	X
Region						

Table 4.2: Average Elasticity Estimates

Note: "Demographic" refers to Log-Drivers, Log-Size, Log-Age and Lifecycle Variables

elasticity, where higher weight is given to areas of higher density.[32] If the true model were approximately linear, all these estimates should roughly coincide — for income elasticities this is true, however this is far from true for price elasticities.[33] The average price elasticity estimates are roughly double the OLS coefficients for the 1988 sample, and small and positive for the 1991 sample.

To understand this pattern, consider first the estimated price structure from the 1988 sample, displayed as Figure 5. Here we see downward sloping areas over areas of higher price density, which is what the average derivative estimates are saying (the OLS elasticities are smaller, because they are based on measuring one slope through the two downward sloping areas), with a structure roughly

[32]If f^* denotes the log-income, log-price density, then the density weighted average derivatives are $\delta_{f^*,LY} \equiv E\left[f^* \frac{\partial G^*}{\partial LY}\right]/E[f^*]$; $\delta_{f^*,LP} \equiv E\left[f^* \frac{\partial G^*}{\partial LP}\right]/E[f^*]$. All average derivatives are estimated nonparametrically using the instrumental variables methods discussed in Stoker (1992, pp. 61-63).

[33]Table A.4 gives average derivative estimates for log-income and log-age from the partial linear model (3.1) of the last section. Here there is no erratic variation of the coefficient estimates, as one would expect from the nearly linear structure we discovered there.

comparable to the Hausman-Newey plot earlier.[34] However, we can also see where heterogeneity exists that has not been accounted for. In particular, Figure 5 shows mainly low price and high price modes, with few price values in between.

Where might one see a bimodal distribution for gasoline prices? In any gas station, under the headings "Regular" and "Premium." While totally obvious in retrospect, it is clear that modeling different types of gasoline might help explain the differential price reactions. Figure 6 contains an analogous plot for 1991; here the price distribution is somewhat more even, and the nonlinearities less pronounced, but this still confounded by the presence of different gasoline types.

We found that households were asked whether they bought regular, premium or both kinds of gasoline, and set out to put the price measures on the same footing. We tried various methods of estimating a regular price level for all households and including it in the regression analysis (so that we would measure the impact of a proportional change in all prices). We obtained the same results as another method, namely just restricting attention to households who purchased regular gasoline only. Table 4.3 contains the estimates for the 1445 "regular only" households for 1991. These estimates give much higher price elasticities than before; in fact, they are so high as to cause further skepticism. Examining the plot of price structure for these households confirms the further problems. In Figure 7, we can see that there is essentially there is only an effect of observed prices starting in the very top of the price range — and that effect is so strong as to cause the huge estimates of the price elasticity.

Of course, there are several ways to look further into this — for instance, the choice of "regular" is actually endogenous, and therefore gasoline type must be used with care.[35] However, we used yet another tool of analysis: the telephone. We called the Energy Information Agency to find out exactly how the expenditure data were collected. We learned that the EIA compiled average prices for regular and premium gasoline for each region for each month. Each survey household was asked for what months it owned each of its vehicles and whether each vehicle used regular gasoline, premium gasoline, or about the same amounts of each.[36]

[34] Figure 5 is drawn with log-price, the regressor variable, on the horizontal axis, in contrast with the figures from Hausman and Newey (forthcoming), which plot log-price on the vertical axis.

[35] Premium gasoline is primarily sold for use in automobiles that would exhibit very similar performance with regular gasoline.

[36] No other answers were allowed, except that in 1991 but not 1988, households were asked whether they used unleaded or leaded regular gasoline, and different average prices were em-

LTGALS Dep. Var. 1445 Obs.	1991 OLS	ADE
LP	-1.40	-3.80
	(.539)	(2.25)
LY	.127	.151
	(.019)	(.024)
Demographic	X	X
Residence	X	X
Region		

Table 4.3: Price Elasticities, Households Using Regular Gasoline
Note: "Demographic" refers to Log-Drivers, Log-Size, Log-Age and Lifecycle Variables

Average fuel costs for each vehicle were then computed as the weighted average of the appropriate monthly prices, where the weights were national average miles driven in each month.

Thus all consumers in each region were *assumed* to face the same prices. Variations in average cost and thus in LP reflected households' choice of fuel type and changes in vehicle ownership during the year. Intra-regional differences in average cost in the RTECS data do not measure differences in prices faced by households, and inter-regional differences are completely accounted for by our regional dummy variables. Having learned this, we decided against trying to estimate price elasticities. Had we not used nonparametric techniques in this study, however, we would very likely have simply reported the "reasonable" elasticities estimated when regional variables are excluded.[37]

ployed depending on the answer.

[37]The question of how to interpret the patterns in Figures 6 and 7 is open to conjecture. One possibility is that the highest prices arise in the winter months in northern states, so that the pattern is actually a depiction of seasonal driving habits (little driving for cars owned only in the winter).

5. Concluding Remarks

We have discovered that household demographic structure has strong effects on gasoline demand. The most striking is how the inclusion of the number of drivers cuts the estimate of income elasticity by half. This finding suggests why gasoline demand grew rapidly in the U.S. in recent decades and why future growth may not be so robust.

We found a separate effect of the number of household members, beyond those licensed to drive. On the question of whether rich households taper off in their use of gasoline, we have found no such effect. We discovered only that at low incomes, the income elasticity is zero, possibly because there is a subsistence gasoline consumption level in the U.S. We found that there is no age effect until the age-of head is roughly 50, whence a fairly sharp decrease with age commences. In all specifications, we verified that urban households drive less than suburban households, who drive less than rural households. We verified the basic model with a battery of specification tests, including comparisons of the 1991 data with 1988 data, as well as differenced estimates from a small household panel. Finally, we described the erratic structure produced by using price data from the RTECS survey, which led to our discovery of serious problems in the construction of those data.

We have made use of various semiparametric and nonparametric methods to guide our analysis. We feel that the figures depicting the income-age structure, and later the apparent price structure, show the basic data relationships in a much more convincing fashion that the results of a specification search with just OLS regression methods.[38] This application also displays many typical features of econometric work, namely many regressors, many of which are discrete, so it is likely that the methods we have used will be applicable in many other contexts. As such, our work can be viewed as giving illustration of semiparametric and nonparametric methods to two different kinds of problems; namely ascertaining functional specification (the income-age structure) and model diagnostics (the difficulties with price effects).

We have not discussed the basic goodness-of-fit of our equations. For cross section analysis, the degree of fit ($R^2 \cong .40$) of our basic equations is quite good.

[38] We did not uncover the final specifications ourselves by guessing variables to enter in OLS regressions (we tried!) before applying the semiparametric methods. But of course, it is possible that cleverer researchers could have discovered the structure without recourse to those methods.

But here it is important to remember how the basic modeling framework is central to the interpretation of the results. For instance, it has been suggested to us that we attempt to control for the "driving technology" in our regressions. In Table A.5, we present the results from including log-number of cars owned and log-miles-per-gallon in this spirit, and we find that the fit of the equation has increased substantially ($R^2 \cong .60$). However, number of cars, miles-per-gallon, and other elements of "driving technology" are clearly endogenous to the decision framework, with all coefficient estimates subject to familiar biases. Moreover, we have not been able to come up with instruments, or equivalently, any observable features that would be associated with differences in number of cars but not also associated with gasoline consumption. In any case, it is important to keep in mind that the modeling framework is the essential precursor to any statistical analysis, even for studying a problem as straightforward as the household demand for gasoline.[39]

References

[1] Aitcheson, J. and J.A.C. Brown (1963), *The Lognormal Distribution.* Cambridge (UK): Cambridge University Press.

[2] Aït-Sahalia, Y., P. Bickel and T.M. Stoker (1994), "Goodness-of-Fit Tests for Regression Using Kernel Methods," MIT Sloan School of Management Working Paper No. 3-747, November.

[39] A stark, but somewhat silly illustration of the endogeneity problem is as follows. Consider the following regression:

$$LTGALS \cong \underset{(.009)}{.91}\, LMILES - \underset{(.007)}{.03}\, LY + \underset{(.018)}{.02}\, LDRVRS + \underset{(.013)}{.06}\, LSIZE + \underset{(.015)}{.09}\, LAGE$$

for 1991, that has $R^2 = .87$.

It seems natural to make the assumption that annual MPG experienced by a household is not related to how many miles they drive in a year. But then the identity

$$LTGALS = LMILES - \ln \text{MPG}$$

says that $LMILES$ has a true coefficient of 1 in an $LTGALS$ equation. Consequently, the above regression illustrates both a dramatic improvement in fit, as well as very biased coefficients, due to the endogeneity of $LMILES$.

[3] British Petroleum Company (1993), *BP Statistical Review of World Energy.* London: The British Petroleum Company, June.

[4] Chui, C.K. (1988), *Multivariate Splines*, CBMS-NSF Series in Applied Mathematics #54, Philadelphia, SIAM Publishing.

[5] Dahl, C. (1993), "A Survey of Energy Demand Elasticities in Support of the Development of the NEMS," mimeographed, Department of Mineral Economics, Colorado School of Mines, October.

[6] Dahl, C. and T. Sterner (1991), "Analyzing Gasoline Demand Elasticities: A Survey," *Energy Economics*, July, 203-210.

[7] Ellerman, A.D. (1993), "Twenty-Five Years Since the First Oil Shock: What Has Changed?" Talk presented at the Fall Workshop of the MIT Center for Energy and Environmental Policy Research, November 19, 1993.

[8] Energy Information Administration (EIA) (1990), *Household Vehicles Energy Consumption 1988*, Washington: Energy Information Administration, DOE/EIA-0464(88), February.

[9] Energy Information Administration (EIA) (1993), *Household Vehicles Energy Consumption 1991*, Washington: Energy Information Administration, DOE/EIA-0464(91), December.

[10] Engle, R.F., C.W.J. Granger, J. Rice and A. Weiss (1986), "Semiparametric Estimates of the Relationship Between Weather and Electricity Sales," *Journal of the American Statistical Association*, 81, 310-320.

[11] Federal Energy Administration (FEA) (1974), *Project Independence Report*, Washington: Government Printing Office, November.

[12] Gately, D. (1990), "The U.S. Demand for Highway Travel and Motor Fuel," *The Energy Journal*, 11, July, 59-73.

[13] Härdle, W. (1991) *Applied Nonparametric Regression*, Cambridge (UK): Cambridge University Press (Econometric Society Monographs).

[14] Hausman, J. A. (1975), "*Project Independence Report*: An Appraisal of U.S. Energy Needs Up to 1985," *Bell Journal of Economics*, 6, Autumn, 517-551.

[15] Hausman, J.A. and W.K. Newey (forthcoming), "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," forthcoming in *Econometrica.*

[16] Jorgenson, D, W., D.T. Sleznick and T.M. Stoker (1988), "Two-Stage Budgeting and Exact Aggregation," *Journal of Business and Economic Statistics*, 6, July, 313-325.

[17] Nordhaus, W.D. (1994), *Managing the Global Commons*, Cambridge, MIT Press.

[18] Porter, E.D. and G. Prasad Rao (1993), "Demand for Highway Fuels in the U.S.: Trends, Prospects and Major Uncertainties," Washington: American Petroleum Institute, Draft Working Paper, August.

[19] Robinson, P.M. (1988), "Root-N Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.

[20] Schmalensee, R., T.M. Stoker and R.A. Judson (1995), "World Energy Consumption and Carbon Dioxide Emissions: 1950-2050," Working Paper no. 95-001, Cambridge, MIT Center for Energy and Environmental Research, revised June.

[21] Silverman, B.W. (1986), *Density Estimation*, London, Chapman and Hall.

[22] Stoker, T.M. (1992), *Lectures on Semiparametric Econometrics*, Louvain-la-Neuvre (Belgium), CORE.

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

[24] World Bank (1994), *World Development Report 1994*. Oxford: Oxford University Press.

A. Additional Tables

Variable	1991		1988	
	Mean	St. Deviation	Mean	St. Deviation
LTGALS	6.761	.691	6.800	.672
LY	3.324	.801	3.245	.774
LDRVRS	.550	.399	.6125	.409
LSIZE	.884	.535	.896	.523
LAGE	3.761	.363	3.751	.366
URBAN	.285	.452	.285	.452
SUBURBAN	.444	.497	.474	.499
RURAL	.271	.444	.241	.427
REG1	.076	.265	.049	.215
REG2	.128	.335	.147	.354
REG3	.141	.348	.182	.386
REG4	.143	.351	.080	.270
REG5	.118	.322	.155	.362
REG6	.082	.274	.064	.245
REG7	.080	.272	.101	.301
REG8	.084	.277	.062	.241
REG9	.148	.355	.162	.368
LIFE1	.127	.333	.126	.331
LIFE2	.215	.411	.196	.397
LIFE3	.072	.259	.076	.265
LIFE4	.084	.277	.102	.303
LIFE5	.160	.367	.174	.380
LIFE6	.161	.367	.165	.371
LIFE7	.046	.208	.048	.214
LIFE8	.066	.248	.053	.224
LIFE9	.069	.254	.061	.239

Table A.1: Summary Statistics

LTGALS - Dependent Variable	1991 Semiparametric Estimate	s. e.	Basic Model Estimate	s. e.	1988 Semiparametric Estimate	s. e.	Basic Model Estimate	s. e.
URBAN	-.173	.025	-.175	..026	-.123	.025	-.115	.025
RURAL	.105	.027	.100	.027	.099	.027	.084	.027
REG2	-.049	.048	-.043	.047	-.030	.053	-.035	.054
REG3	.087	.047	.082	.047	.063	.052	.063	.052
REG4	.131	.047	.139	.047	.134	.059	.110	.060
REG5	.105	.049	.113	.048	.119	.053	.111	.054
REG6	.151	.053	.144	.053	.154	.062	.136	.062
REG7	.178	.053	.166	.053	.223	.057	.208	.057
REG8	.119	.052	.146	.052	.185	.063	.144	.063
REG9	.053	.046	.040	.046	.088	.053	.080	.054
LIFE2	.064	.041	.080	.040	.030	.042	.043	.042
LIFE3	.043	.056	.051	.055	-.027	.057	-.004	.055
LIFE4	.046	.055	.049	.054	-.039	.054	-.052	.054
LIFE5	.111	.055	.127	.053	.002	.055	.041	.055
LIFE6	.161	.076	.249	.071	-.194	.076	-.038	.073
LIFE7	.277	.083	.272	.082	.126	.080	.123	.081
LIFE8	.128	.083	.116	.083	.043	.083	.070	.083
LIFE9	.092	.101	.184	.098	-.231	.100	-.136	.099
Constant	Level of $\hat{G}_{1991}$		6.08	.089	Level of $\hat{G}_{1988}$		6.19	.091

Table A.2: Coefficients of Lifecycle and Location Variables

LTGALS - Dependent Variable	Semiparametric Model	OLS Estimates	
LY12-		.053	
		(.044)	
LY12+		.165	.174
	$\hat{G}(LY, LAGE)$	(.020)	(.019)
	(for 1988)		
LAGE50-		.016	
		(.077)	
LAGE50+		-.726	-.712
		(.175)	(.169)
LDRVRS	.653	.668	.672
	(046)	(.045)	(.045)
LSIZE	.064	.061	.058
	(054)	(.054)	(.054)
Residence	X	X	X
Region	X	X	X
Lifecycle	X	X	X

Table A.3: Principal Coefficient Estimates: 1988

LTGALS Dep. Var.		1991	
	OLS	ADE	ADE (dw)
LY	.175	.206	.212
	(.015)	(.020)	(.023)
LAGE	-.283	-.421	-.377
	(.061)	(.046)	(.047)
Demographic	X	X	X
Residence	X	X	X
Region	X	X	X

Table A.4: Average Derivative Estimates For The Basic Model

LTGALS	1991		
Dep. Var.		s.e.	t-stat.
LCARS	.805	(.024)	(33.2)
LMPG	-.394	(.033)	(11.6)
LY12-	.010	(.035)	(.273)
LY12+	.101	(.016)	(6.03)
LAGE50+	-1.07	(.134)	(-8.01)
LDRVRS	.181	(.036)	(4.91)
LSIZE	.085	(.043)	(1.98)
URBAN	-.113	(.021)	(-5.40)
RURAL	.034	(.022)	(1.56)
Lifecycle	X		
Region	X		
R^2	.604		
N	2684		

Table A.5: Regression Including Number of Cars Owned

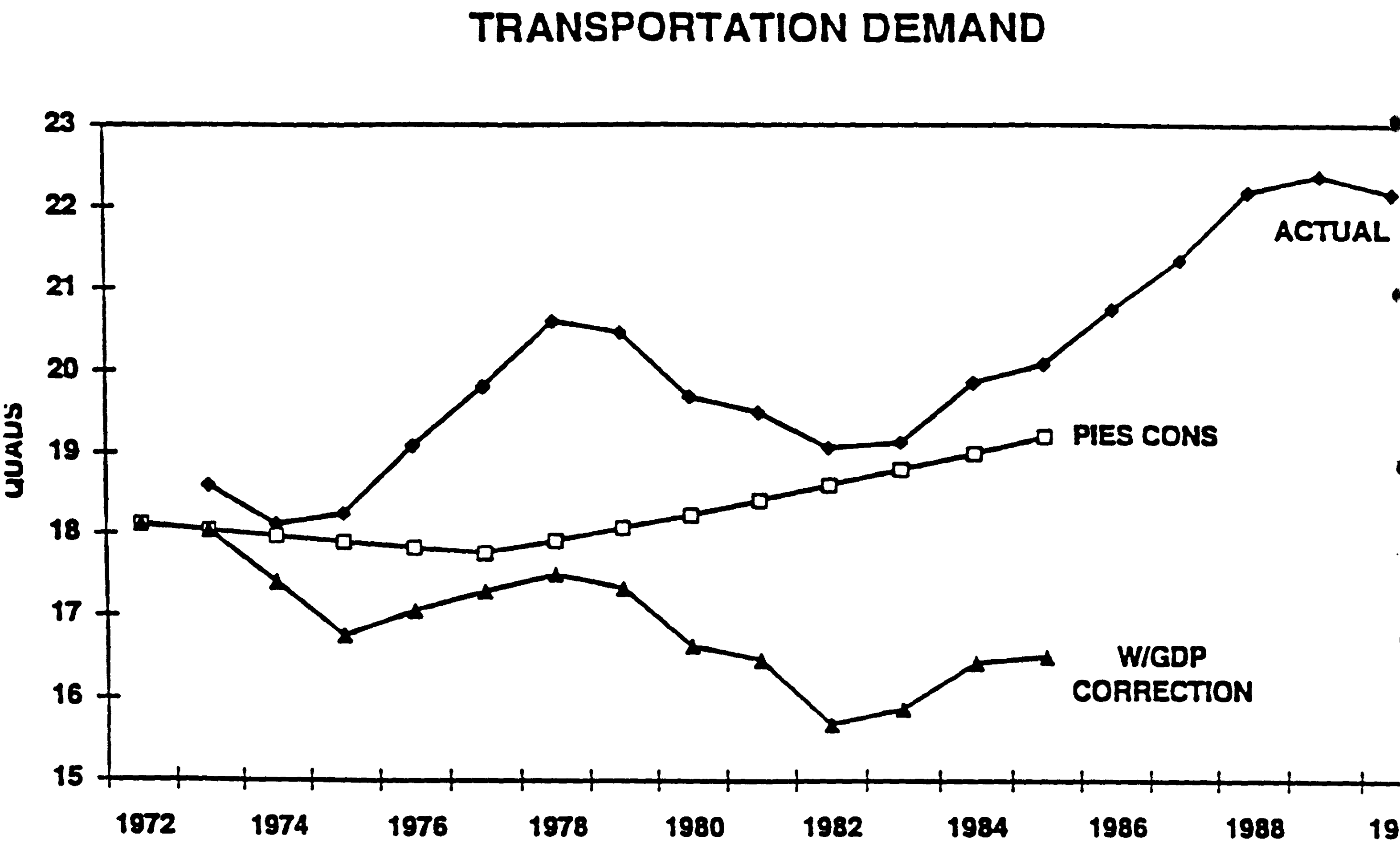


Figure 1
(Source: Ellerman (1995))

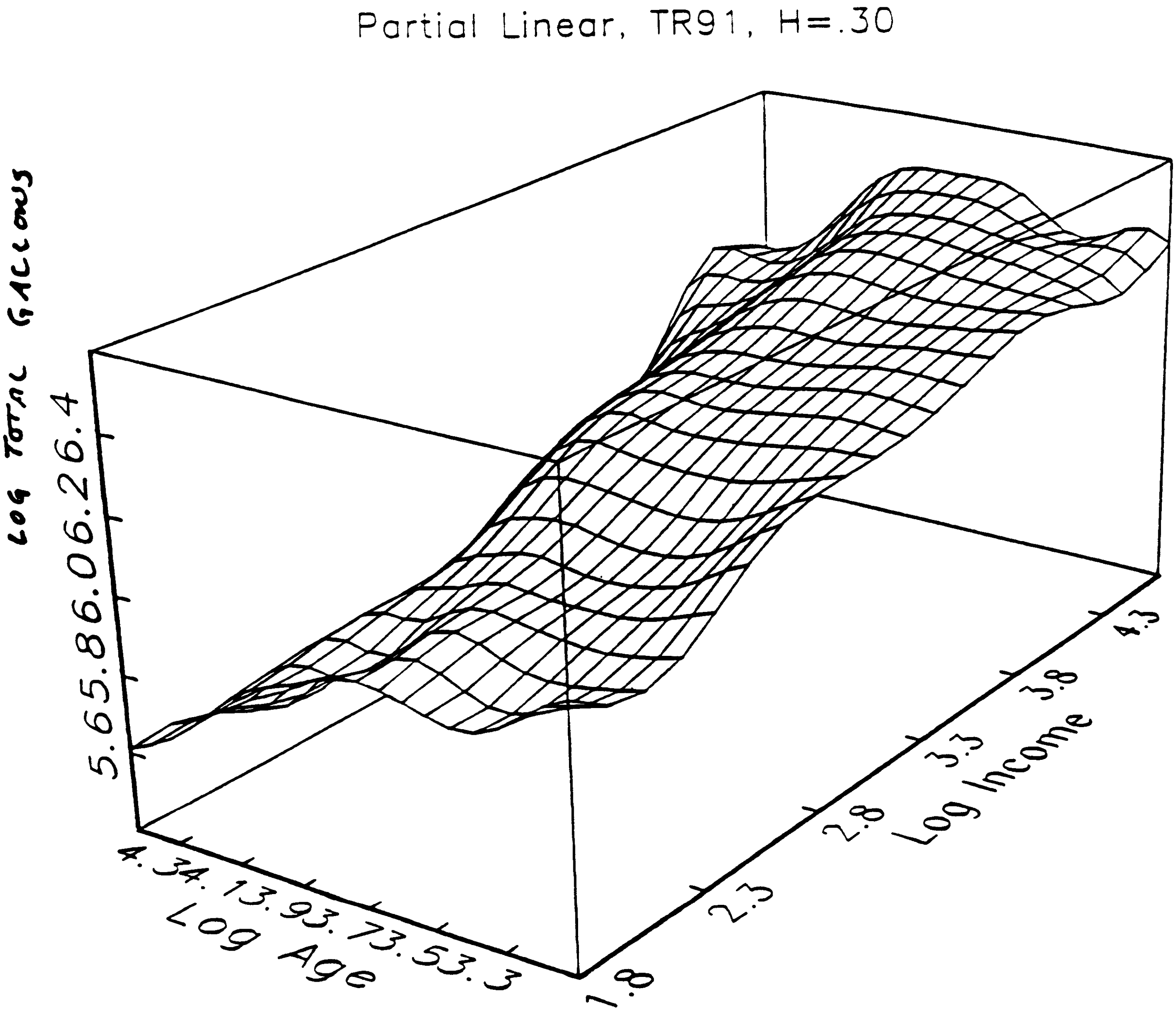

Partial Linear, TR91, H=.30
LOG TOTAL GALLONS
5.65.86.06.26.4
4.34.13.93.73.53.3
Log Age
1.8
2.3
2.8
3.3
3.8
4.3
Log Income

Figure 2a

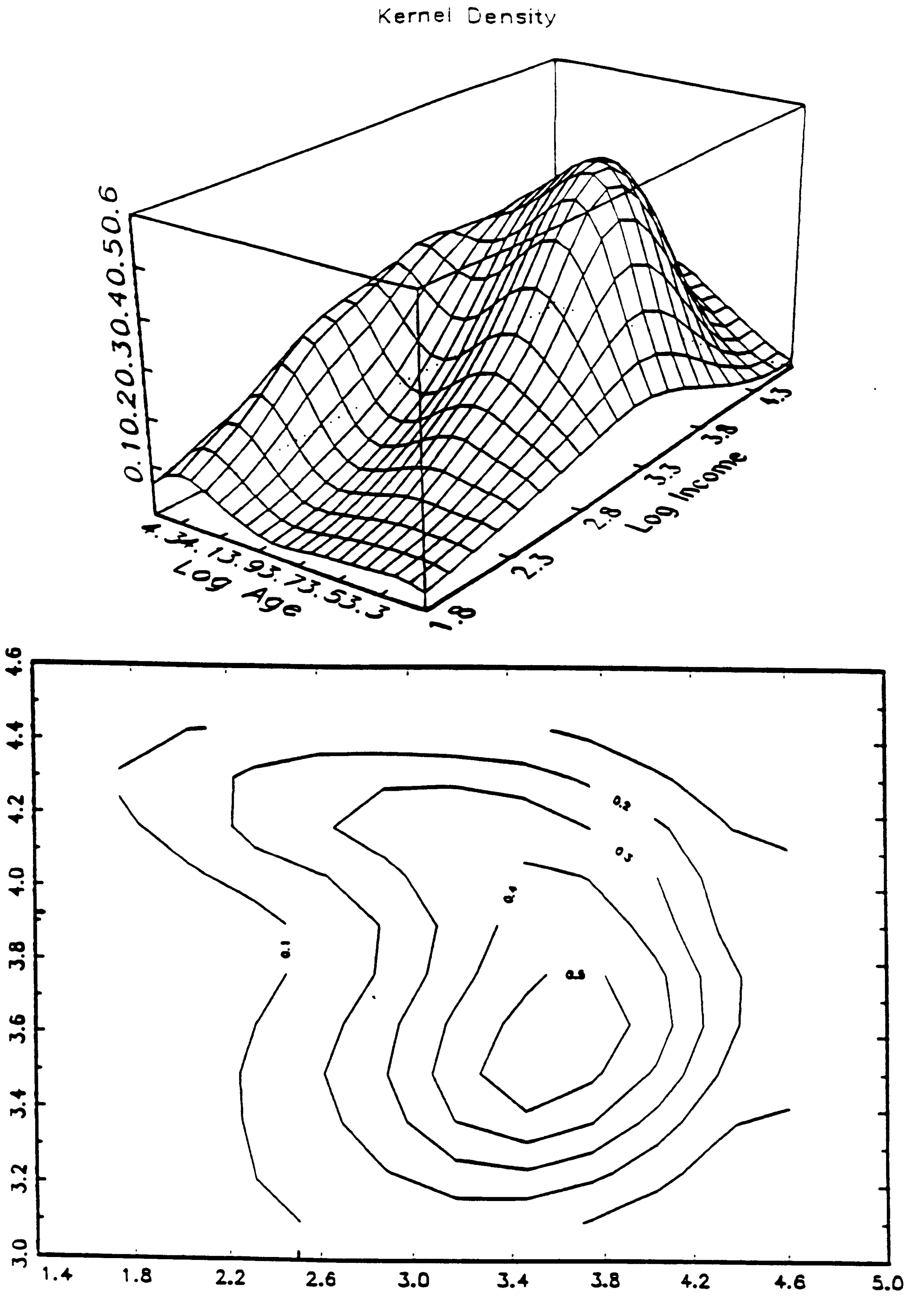
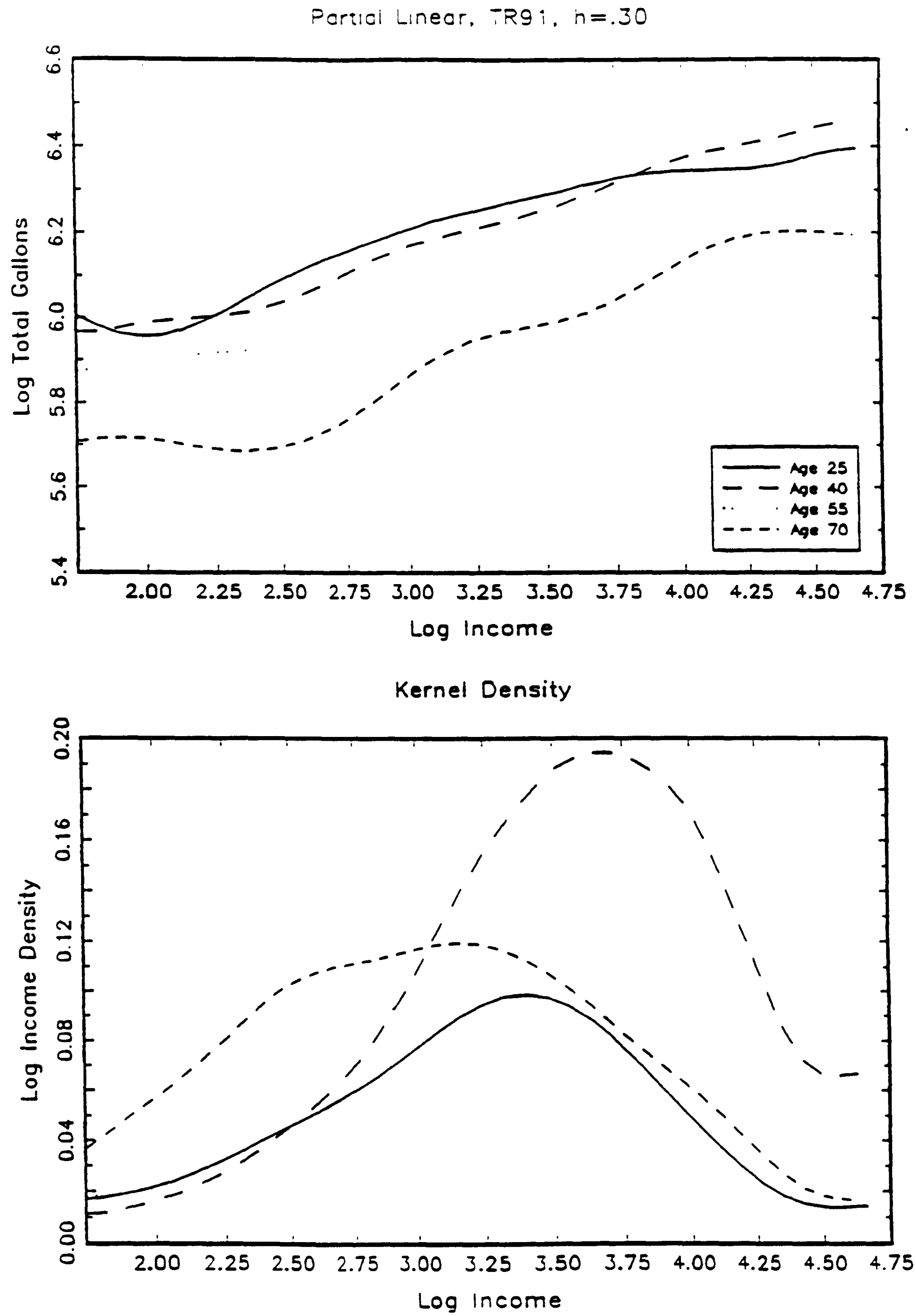
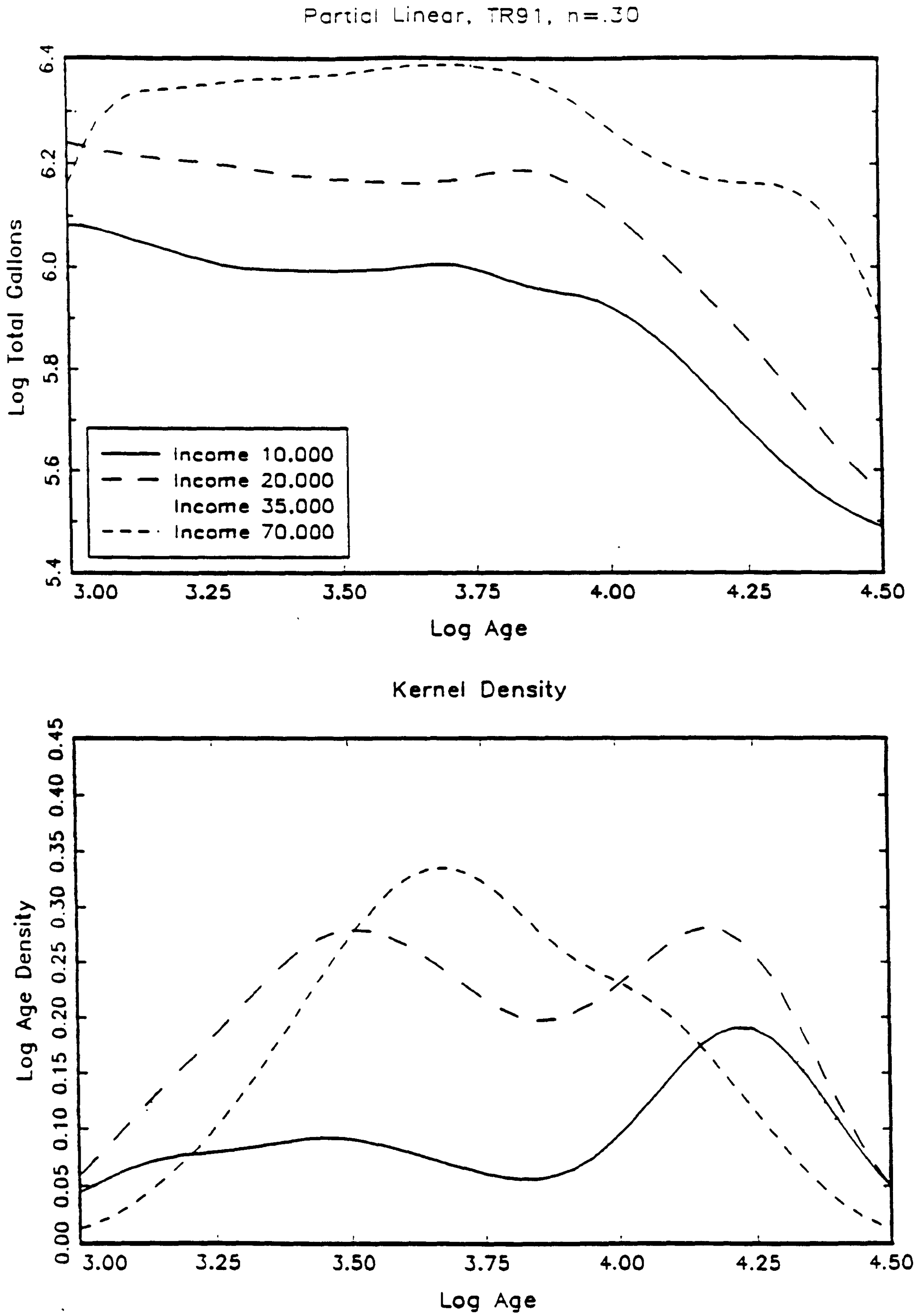

Kernel Density
0.10.20.30.40.50.6
4.34.13.93.73.53.3
Log Age
1.8
2.3
2.8
3.3
3.8
4.3
Log Income
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
0.1
0.2
0.3
0.4
0.5
1.4
1.8
2.2
2.6
3.0
3.4
3.8
4.2
4.6
5.0

Figure 2b



Partial Linear, TR91, h=.30
Log Total Gallons
6.6
6.4
6.2
6.0
5.8
5.6
5.4
2.00
2.25
2.50
2.75
3.00
3.25
3.50
3.75
4.00
4.25
4.50
4.75
Log Income
Age 25
Age 40
Age 55
Age 70
Kernel Density
Log Income Density
0.20
0.16
0.12
0.08
0.04
0.00
Log Income

Figure 3a



Partial Linear, TR91, n=.30
Log Total Gallons
5.4
5.6
5.8
6.0
6.2
6.4
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Log Age
Income 10,000
Income 20,000
Income 35,000
Income 70,000
Kernel Density
Log Age Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40
0.45
3.00
3.25
3.50
3.75
4.00
4.25
4.50
Log Age

Figure 3b

Kernel Regression

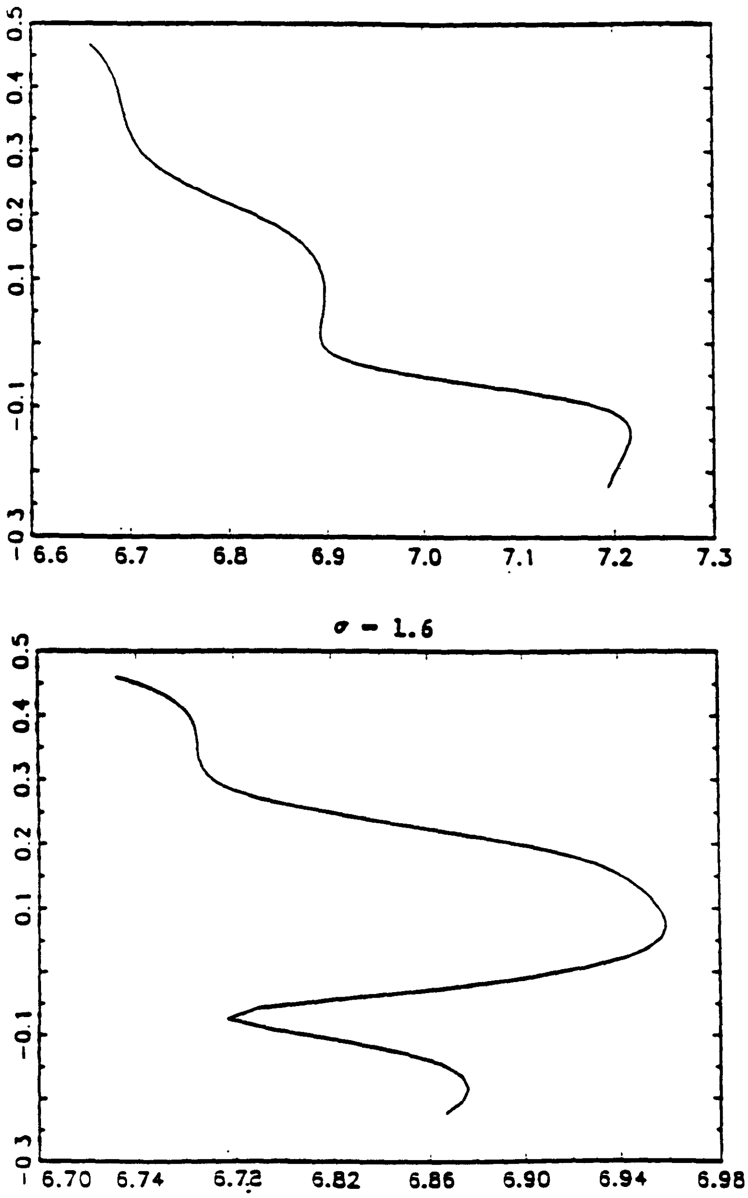


Figure 3

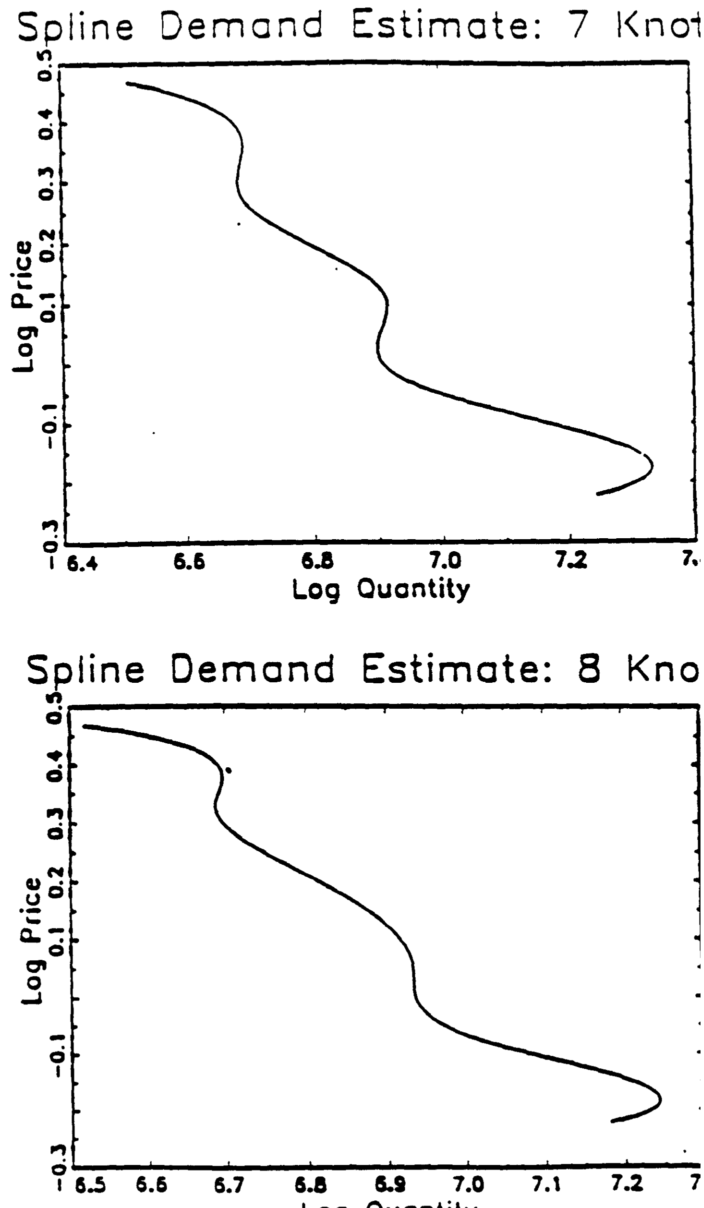


Figure 4

Figure 4

(Source: Hausman and Newey (Forthcoming))

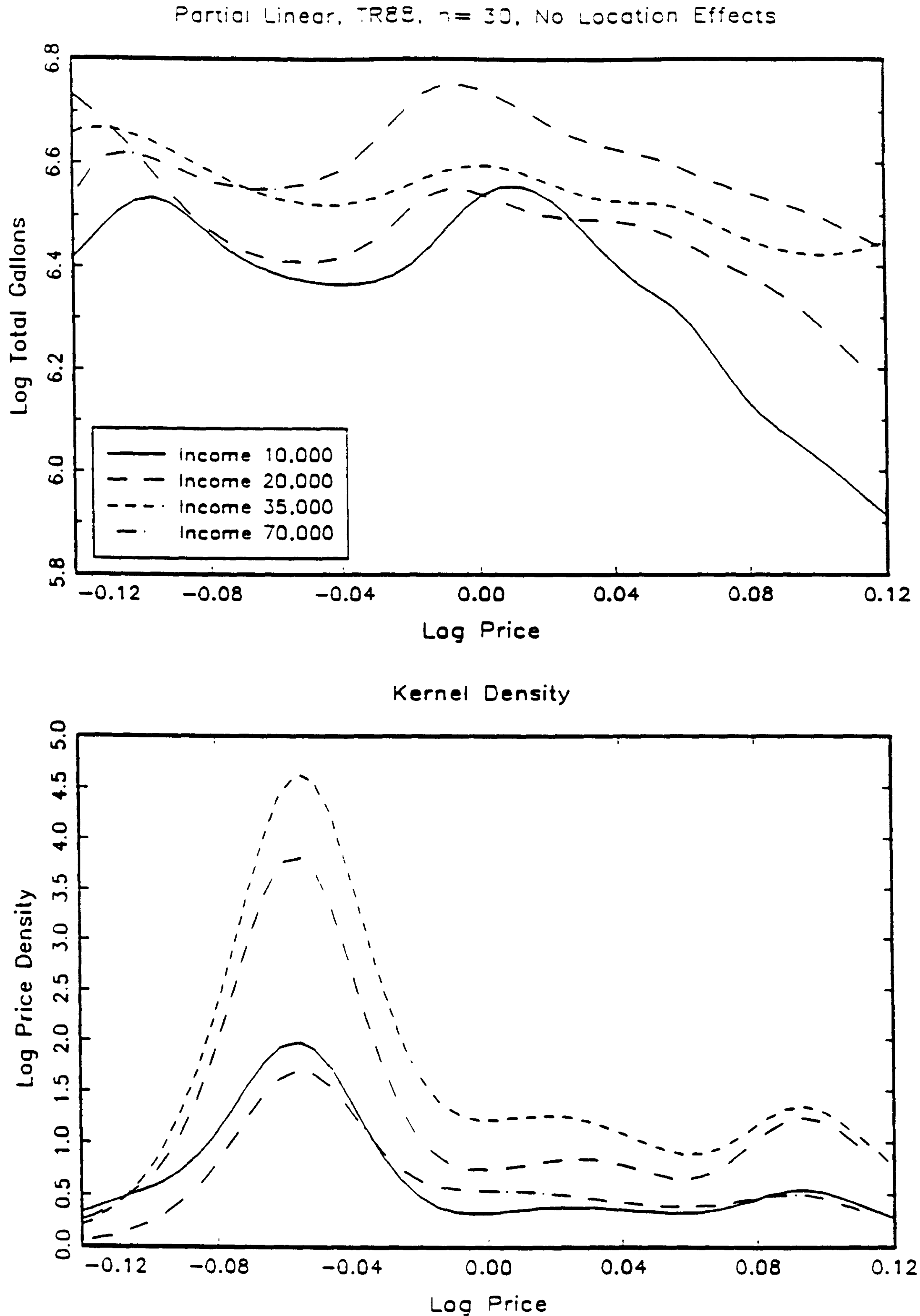

Partial Linear, TR88, h= 30, No Location Effects
Log Total Gallons
6.8
6.6
6.4
6.2
6.0
5.8
Income 10,000
Income 20,000
Income 35,000
Income 70,000
-0.12
-0.08
-0.04
0.00
0.04
0.08
0.12
Log Price
Kernel Density
Log Price Density
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-0.12
-0.08
-0.04
0.00
0.04
0.08
0.12
Log Price

Figure 5

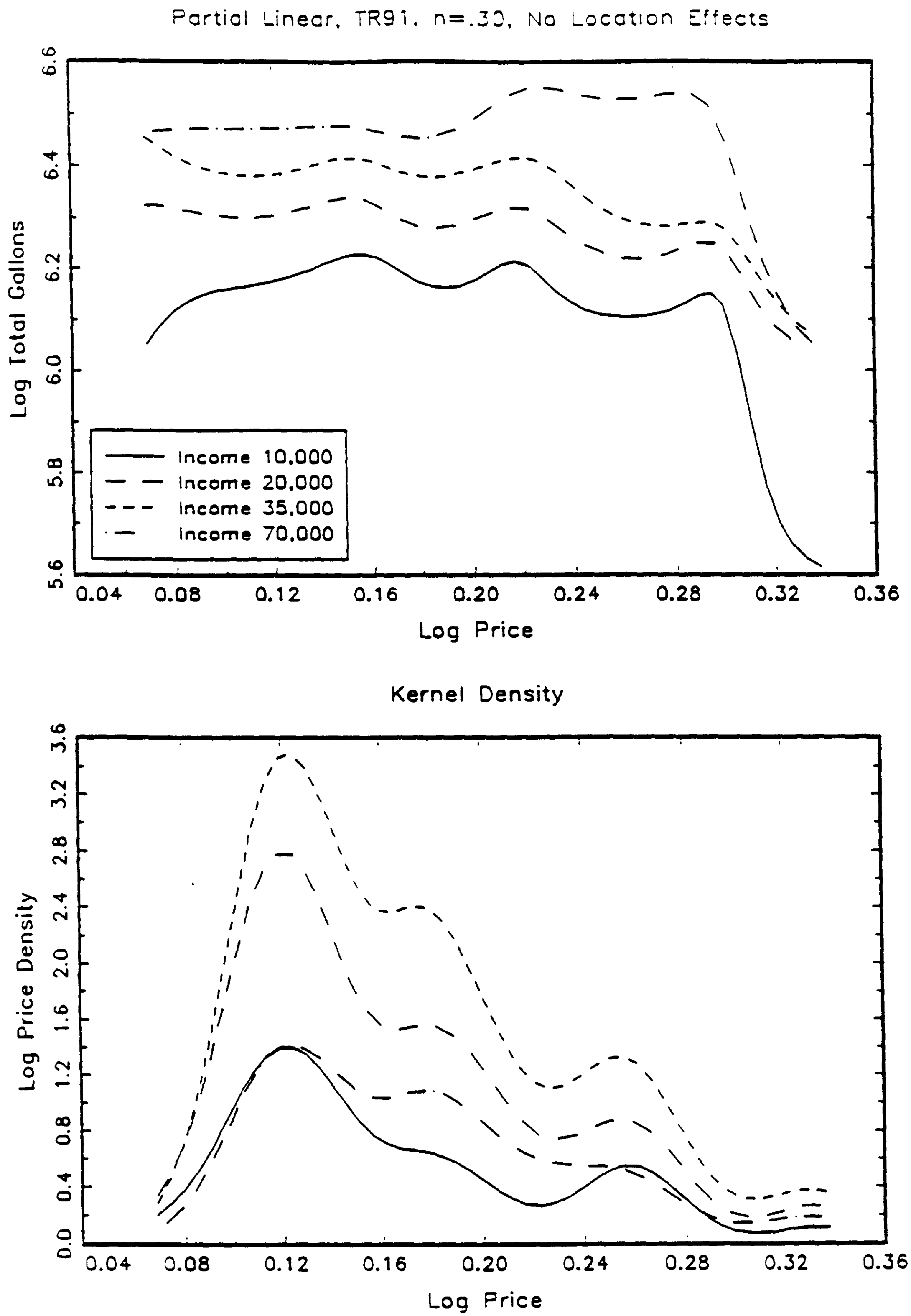
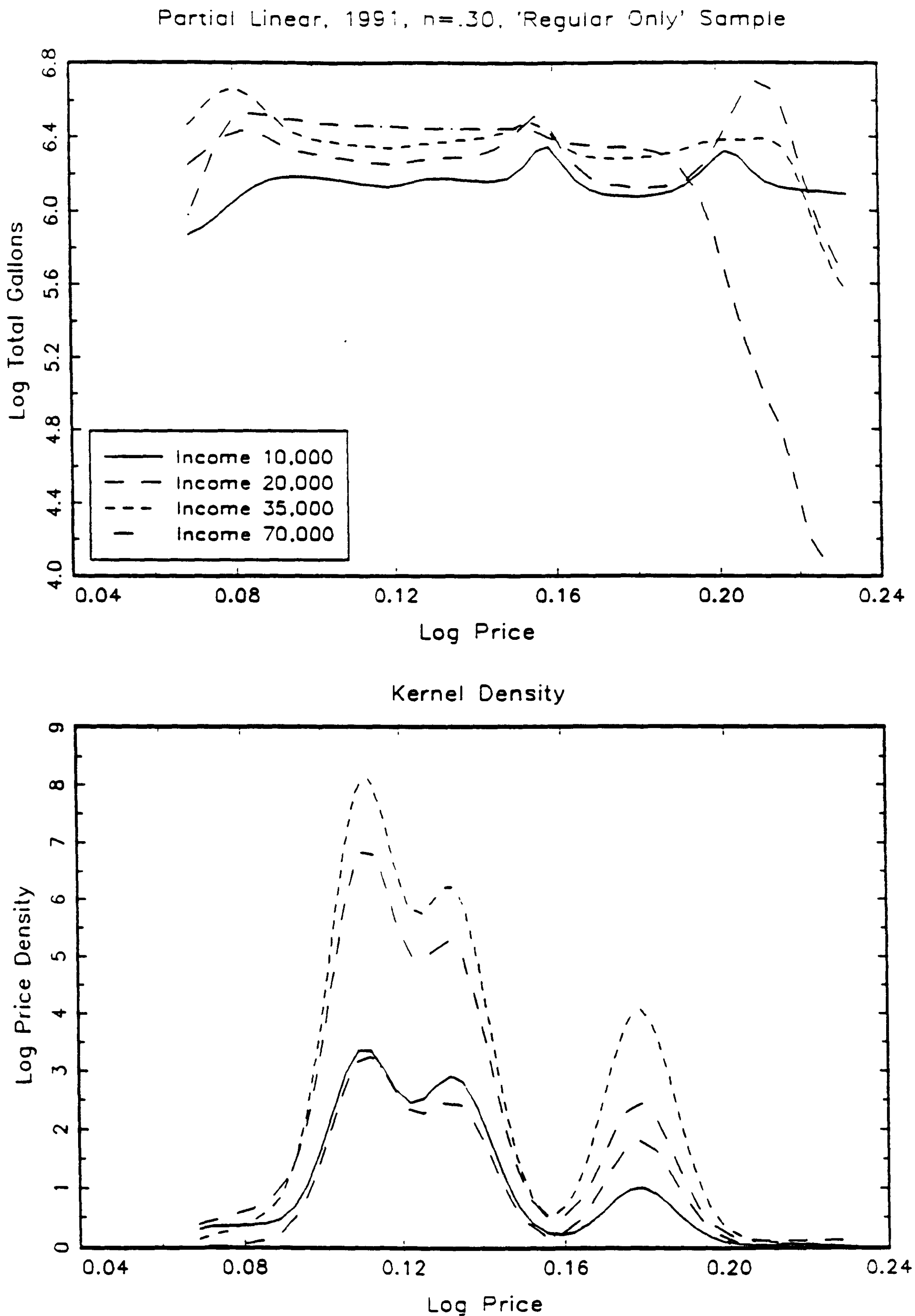

Partial Linear, TR91, h=.30, No Location Effects
Log Total Gallons
6.6
6.4
6.2
6.0
5.8
5.6
Income 10,000
Income 20,000
Income 35,000
Income 70,000
0.04
0.08
0.12
0.16
0.20
0.24
0.28
0.32
0.36
Log Price
Kernel Density
Log Price Density
3.6
3.2
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0
0.04
0.08
0.12
0.16
0.20
0.24
0.28
0.32
0.36
Log Price

Figure 6



Partial Linear, 1991, n=.30, 'Regular Only' Sample
Log Total Gallons
Log Price
Income 10,000
Income 20,000
Income 35,000
Income 70,000
Kernel Density
Log Price Density
Log Price

Figure 7