Econometrica, Vol. 69, No. 6 (November, 2001), 1697–1709

HOUSEHOLD GASOLINE DEMAND IN CANADA

By Adonis Yatchew and Joungyeo Angela No[1]

1. introduction

Several recent studies analyze demand for gasoline in the U.S. at the level of the household.[2] Using data from six surveys conducted between 1979 and 1988 Hausman and Newey (1995, 1998) assemble a massive data set of over 18,000 observations and estimate a partial linear model that pays careful attention to price and income effects. The respective elasticities are about $-.8$ and .37. The authors also conclude that welfare effects of tax changes are substantially different if one estimates price effects nonparametrically rather than parametrically.

Schmalensee and Stoker (1999) estimate a similar model using 1988 and 1991 data (with approximately 5000 observations), but pay careful attention to the modeling of demographic effects. They conclude that variables such as the number of licensed drivers and age have very substantial impacts and should be incorporated in forecasts of gasoline demand. The inclusion of demographic effects cuts the income elasticity in half to about .21. They also find that the price variable available in their data is unreliable because it is imputed by the Energy Information Administration rather than actually collected at the household level. The authors attempt to control for price variation by the inclusion of regional dummy variables. However, the lack of price data at the household level precludes the estimation of a price elasticity.

The absence of demographic effects in the Hausman and Newey model and the unavailability of price data in the Schmalensee and Stoker work raises the question whether the inclusion of such data would substantially alter their respective conclusions. Fortunately, recent Canadian data are available that contain good price *and* demographic detail. We estimate a model that is similar to our predecessors. The estimates reveal strong demographic effects—in particular the number of licensed drivers and age—that are numerically consistent with the findings of Schmalensee and Stoker. Our estimated price elasticity is close to that found by Hausman and Newey and our income elasticity lies between the two papers. Not surprisingly, we find that price is essentially orthogonal to demographic variables, thus the absence of demographics in Hausman and Newey and the absence of reliable price data in Schmalensee and Stoker would appear not to compromise either set of results. We also estimate separate models for regular, medium, and premium grades. Although nonparametric price effects—particularly for premium gasoline—display patterns similar to those found by Hausman and Newey, our data do not reject a log-linear specification for the price effect.

Finally, we consider the possibility that price is endogenous. It is not unusual for the coefficient of variation of the price of regular gasoline to be as high as 5% *within* a given urban area (this is more than half the coefficient of variation for our whole data

[1] The authors are grateful to a co-editor, Campbell Watkins, and two anonymous referees for helpful comments.

[2] There is also a large literature on demand for gasoline that uses aggregate data. For recent surveys see, e.g., Espey (1996, 1998).

set). Thus, individuals who drive more are likely to encounter a broader range of prices and hence incur lower search and transaction costs. Furthermore, these same individuals derive greater benefit from cheap gas and therefore would be willing to accept higher search and transaction costs. One might therefore expect the actual price paid to be negatively correlated with the residual in an equation where the dependent variable is distance traveled or gasoline consumed. This in turn would lead to overestimation of the (absolute) price elasticity. In the concluding section we conduct a simple test of the endogeneity hypothesis.

Our basic specification is the partial linear model with price and age entering nonparametrically. We rely principally on differencing techniques (Yatchew (1997, 1999)) that substantially simplify estimation and testing. As a check, we also employ Robinson's (1988) double residual method and specification testing procedures of Li (1994), Fan and Li (1996), Zheng (1996), and Aït-Sahalia, Bickel, and Stoker (1998). These are found to yield very similar estimates and conclusions.

In summary, we have four main objectives: to estimate a semiparametric model of gasoline demand using recent Canadian data; to determine whether the inclusion of both price and demographic variables in the same model would likely have altered the Hausman and Newey or Schmalensee and Stoker results; to determine whether separate estimation by grade of gasoline yields different estimates; and to test whether our price variable is endogenous.

2. MODEL AND DATA

Our data are obtained from the National Private Vehicle Use Survey, conducted by Statistics Canada between October 1994 and September 1996. It is a micro-data file that contains household-based information on vehicles, fuel consumption patterns, and demographics. We focus our study on households with a nonzero number of licensed drivers, vehicles, and distance driven. The resulting data set is comprised of 6230 observations.

Data on fuel consumption patterns were obtained from fuel purchase diaries that were kept in the vehicle for a one month period. At each purchase, recorded information included the odometer reading, type of fuel (regular/medium/premium), price per liter, liters purchased, and total paid for fuel. Our price variable was obtained by dividing total expenditures on fuel by the number of liters purchased during the diary month. Given the nature of the data collection mechanism, one would not expect serious errors in variables problems in the price data.

For households with multiple vehicles, the vehicle for which the diary was kept was selected randomly by computer during the interview process. To estimate the total distance driven by each household, we multiply the kilometers driven for the selected vehicle by the number of vehicles in the household.[3] About 56% of households own one car, 38% own two.

[3] Consider, for example, two-vehicle households. Let y_1 and y_2 be distance driven by each of the two vehicles where y_1 corresponds to the vehicle for which the diary is kept. We are interested in how total distance driven y_1+y_2 relates to a vector of conditioning variables x. Since the order of the vehicles is random, we have $E(y_1|x)=E(y_2|x)=1/2E(y_1+y_2|x)$. Thus, multiplying distance traveled by the number of vehicles preserves consistent estimation of conditional means. To ease comparability of elasticities with Hausman and Newey, and Schmalensee and Stoker, we actually take log(number of vehicles $*$ distance traveled by selected vehicle) as our dependent variable.

Data on household income, which refers to the previous year's combined annual household income before taxes, is reported in 9 ranges.[4] We take the midpoint of each interval. For the highest income category (income over \$80,000) we use the conditional median for national household income above \$80,000, which is approximately \$100,000.

Several demographic effects are included. The age variable corresponds to the first member on the household roster that drives the selected vehicle. Actual age is observed to age 65 at which point we know only that the person is over 65. We also include in our model the number of members in the household and the number of licensed drivers.

Since distance driven typically exhibits strong seasonal patterns, we include monthly dummies. City dwellers tend to drive considerably less, which warrants the inclusion of an urban dummy. We also have information on the geographic location of the household (Atlantic provinces, Quebec, Ontario, the Prairie provinces, and British Columbia) but interestingly, these variables were immaterial in virtually all models of distance driven or gasoline consumed. Gasoline taxes vary significantly by province, so geographic location *can* be expected to have an important impact through the price effect. (Summary statistics on gasoline prices by region are provided in Section 3.4, Table II.) Some of the variants that we estimated incorporated parametric interactions between parametric and nonparametric variables (e.g., between income and price), but these were generally not significant.

3. EMPIRICAL RESULTS

3.1. *The Schmalensee and Stoker Specification*

For comparison purposes we begin with a model that comes as close as our data permit to the specification used by Schmalensee and Stoker. Table I compares their 1991 estimates to our estimates for the 1994–96 period where the dependent variable is the log of total monthly gasoline consumption log(GASCON). Our income elasticity is somewhat higher while our age elasticity is lower, but the latter could be due to the fact that our age variable is censored at 65. We will model the age effect more carefully later. The *sum* of the 'drivers' and 'household size' effects (the two variables are correlated) is about .7 in both Canadian and U.S. data. Thus a doubling of household size and number of drivers is predicted to increase miles driven by about 70% in both countries. In the U.S., rural residents drive about 28% more $(.109-(-.172))$ than urban residents, which is very close to the Canadian estimate.[5] Overall, what catches the eye is the striking similarity between parameter estimates despite the fact that these micro-data are from different countries at different points in time.

The omission of the price variable from the Schmalensee and Stoker models could bias their estimates of other effects, though the presence of location dummies would account for inter-regional price variation. To determine whether this omission leads to biases we regressed the log price of gasoline on expanding sets of variables that are potentially relevant for modeling gasoline consumption. Location variables (urban and provincial/regional dummies) explain only about 18% of the variation in price. When grade of gasoline—which Schmalensee and Stoker correctly point out is a critical determinant of price—is

[4] The data set used by Hausman and Newey contains 12 income categories while that of Schmalensee and Stoker has 23 ranges.

[5] In our data, 'urban areas' are towns or cities with population greater 15,000 as of the 1981 Census. In Schmalensee and Stoker, 'urban areas' refer to legally incorporated cities within a Metropolitan Statistical Area (MSA), which in turn contains at least 50,000 inhabitants and includes both urban and suburban areas. Rural areas are those not located within an MSA.

TABLE I
COMPARISON OF U.S. AND CANADIAN ESTIMATES[a]

Dependent Variable	U.S. Data 1991 log(GASCONS)		Canadian Data 1994–96 log(GASCONS)	
	Coef	SE	Coef	SE
log of income spline (income > \$20K)	0.211	0.018	0.268	0.019
log of age spline (age > 50 years)	−0.913	0.081	−0.642	0.090
log of number of drivers	0.620	0.037	0.537	0.030
log of household size	0.097	0.028	0.167	0.024
Urban dummy	−0.172	0.026	−0.272	0.019
Rural dummy	0.109	0.027	—	—
Single adult, age < 35	0.183	0.056	0.154	0.053
R^2	0.396		0.234	
Sample size	2684		6230	

[a] See Schmalensee and Stoker (1999, p. 654, Table II). Both models also include regional dummies.

added to the model, explanatory power increases to 35%. Adding monthly and trend effects raises R^2 to 44.1%. Thus even after removal of these various effects, there is still substantial variation in price that could lead to bias in demographic parameter estimates if price is omitted from the demand equation. As it turns out, the omission of price does not lead to bias—addition of demographic variables raises R^2 very marginally to 44.4% and coefficient magnitudes in the price regression are immaterial. We also performed tests of independence of price (after removal of location, grade, and temporal effects) and income and demographic variables (age, number of drivers, household size, and a dummy for singles under age 35) and found very modest evidence of a relationship. The finding of orthogonality between the price of gasoline and demographic effects is not surprising since the price of gasoline is determined on world markets with differences in average inter-regional prices principally due to taxes.

At this point it is also worth commenting on the implications for the Hausman-Newey analysis, which includes price, income, location, and temporal effects, but omits demographics. For the same reasons indicated above, estimating a model that does not condition on demographics should have no material impact on the estimation of the price effect. Indeed in our regressions, omission of demographics did not alter the estimated price effect.

3.2. *Modeling Age and Price Nonparametrically*

The most general specifications that we consider are the following:

$$y = f(\text{PRICE, AGE}) + z\beta + \varepsilon \tag{3.1}$$

and

$$y = f(\text{PRICE}) + \delta\text{AGE DUMMIES} + z\beta + \varepsilon \tag{3.2}$$

where f denotes a generic smooth function, AGE DUMMIES consist of separate dummy variables for each age from 20 to 65 and a dummy for those over age 65, and z consists of income, demographic, location, and temporal variables. Neither (3.1) nor (3.2) is, strictly speaking, nested in the other: (3.2) does not require a smooth age effect as does (3.1) while (3.1) does not assume separability between PRICE and AGE. Throughout, our dependent variable will be the logarithm of total distance traveled in the month, log(DIST). Results for log(GASCON) are similar.

To estimate the parametric effects using differencing, we reorder the data so that the values of the nonparametric variable(s) are 'close.' We then 'difference' the data to 'remove' the effect of the nonparametric variable(s) and run ordinary least squares regressions of the differenced dependent variable on the differenced parametric explanatory variables. For details on these techniques see Yatchew (1997, 1999). We use tenth order differencing throughout the paper. Results for other orders were similar.

Figure 1 contains our analysis of model (3.1). The upper graph illustrates the scatter of data on PRICE and AGE and the path that we take through the data points for purposes of differencing. Estimates of the parametric effects are presented. Having estimated parametric effects, we remove these from the dependent variable and estimate PRICE and AGE effects nonparametrically (see lower graph in Figure 1). Next, we proceed with various significance and specification tests. To test the joint significance of PRICE and AGE, we compare the estimator of the residual variance from (3.1) to the estimator under H_o: $y = z\beta + \varepsilon$. The test statistic whose distribution is standard normal under the null hypothesis yields $(10 \cdot 6230)^{1/2}(.5130 - .4953)/.4953 = 8.92$, which is strongly significant. Next we test the significance of each nonparametric variable individually. A test of H_o: $y = f(\text{AGE}) + z\beta + \varepsilon$ against (3.l) yields $(10 \cdot 6230/2)^{1/2}(.5067 - .4953)/.4953 = 4.06$. A test of H_o: $y = f(\text{PRICE}) + z\beta + \varepsilon$ yields $(10 \cdot 6230/2)^{1/2}(.5044 - .4953)/.4953 = 3.24$. Thus both variables are significant individually and collectively.[6]

Next we explore parametric specifications for the AGE effect while maintaining a nonparametric PRICE effect. The upper graph in Figure 2 displays a scatter-plot of the estimated AGE DUMMIES along with a kernel smooth of these points. The letter 'R' denotes the estimated dummy for those whose age exceeds 65. Evidently, distance driven is highest during the early 20's, declines steadily to age 30 then remains approximately flat to age 50 at which point it begins to decline again. There also seems to be an increase in distance driven by the household during the late 40's. We suspect that this is due to increased driving by teenagers in the household though sufficiently detailed demographic data are not available to establish intra-family patterns of driving.[7] The kernel smooth of the AGE DUMMIES suggests that a spline function may be adequate. We incorporate a cubic B-spline with knots at five year intervals over the range 20 to 65 and a separate knot for those over 65:

$$y = f(\text{PRICE}) + \gamma bs(\text{AGE}) + z\beta + \varepsilon. \tag{3.3}$$

Here $bs(\text{AGE})$ denotes the *S-Plus* function that generates a basis matrix for polynomial B-splines. Thus the only nonparametric effect in (3.3) is that of the PRICE variable. Figure 2 displays the estimated cubic B-spline, which is seen to track the kernel estimate closely. Testing specification (3.3) against (3.1) we obtain $(10 \cdot 6230/2)^{1/2}(.4967 - .4953)/.4953 = 0.50$. Thus, the cubic spline model would appear to be adequate. If in (3.3) we substitute log(AGE), then we obtain $(10 \cdot 6230/2)^{1/2}(.5014 - .4953)/.4953 = 2.17$, suggesting that the log-linear specification oversimplifies the modeling of the AGE effect.

[6] See Yatchew (1997, Proposition 2) for differencing tests of specification or significance when the null is parametric and Yatchew (1999, Proposition 5.5.2) for similar tests when the null is semiparametric.

[7] Supportive of this point is the observation that when performing the same kind of analysis on 1 and 2 person households, the increased driving in the late 40's does not appear.

	coef	se
log income	0.2870	0.0209
log drivers	0.5318	0.0346
log hhsize	0.1224	0.0284
single, age < 35	0.1984	0.0628
Urban dummy	−0.3314	0.0203
monthly effects[b]	—	—
R^2	.2710	
s^2	.4953	
Sample Size	6230	

[a] Dependent variable is log(DIST) Var(dependent variable): .6794. Order of differencing $m = 10$.

[b] For monthly effects see Figure 3.

Upper panel right illustrates ordering of data: data are reordered first by age; then for even ages, data are reordered so that price is decreasing and for odd ages, data are reordered so that price is increasing.

Lower panel right illustrates nonparametric estimate of AGE and PRICE effects after removal of estimated parametric effects.

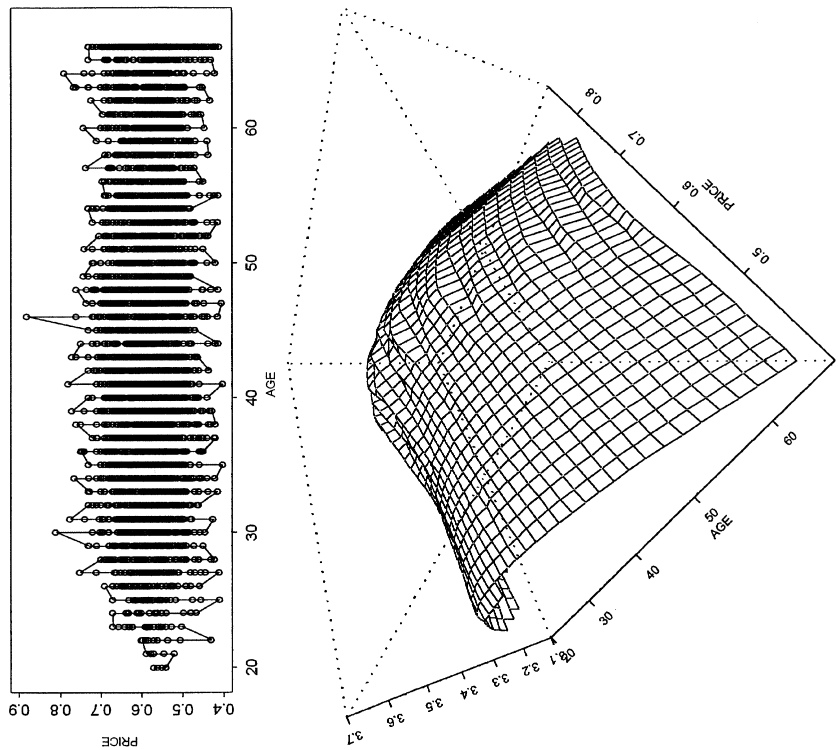


FIGURE 1.—Nonparametric price and age—partial linear model.[a]

Next we explore parametric specifications for PRICE while maintaining a nonparametric AGE effect. A test of H_o: $y = f(\text{AGE}) + \gamma \log(\text{PRICE}) + z\beta + \varepsilon$ against the alternative (3.1) yields $(10 \cdot 6230/2)^{1/2}(.4969 - .4953)/.4953 = 0.57$, suggesting that the log(PRICE) specification should be adequate. The lower graph in Figure 2 illustrates log-linear and nonparametric estimates of the price effect after removal of all parametric and age effects using model (3.2) and (3.3) (the kernel estimates virtually coincide). The log-linear speci-

Specification of Age Effects	Partial Linear Model Nonparametric Price			
	cubic spline		age dummies	
	coef	se	coef	se
log income	0.2844	0.0211	0.2846	0.0213
log drivers	0.5420	0.0345	0.5406	0.0345
log hhsize	0.1101	0.0284	0.1118	0.0284
single, age < 35	0.2054	0.0622	0.2069	0.0627
Urban dummy	−0.3339	0.0202	−0.3324	0.0202
monthy effects[b]	—	—	—	
average price elasticity	−0.890	see below		
R^2	.2689		.2724	
s^2	.4967		.4943	

[a] Dependent variable is log(DIST) Var(dependent variable): .6794. Order of differencing $m = 10$.
[b] For monthly effects, see Figure 3.

AGE EFFECT: scatter plot is individual estimated age dummies, solid line is kernel smooth of these points; dashed line is cubic B-spline with knots at 20,25,30,35,40,45,50,55,60,65.

PRICE EFFECT: solid line is linear effect of log(price)—estimated elasticity is −896 (se = .108); dashed lines are kernel estimates of price effects after estimated parametric effects have been removed using models (3.2) and (3.3); estimated average derivatives from −.870 (se = .101) to −.910 (se = .100) depending on the estimator used.

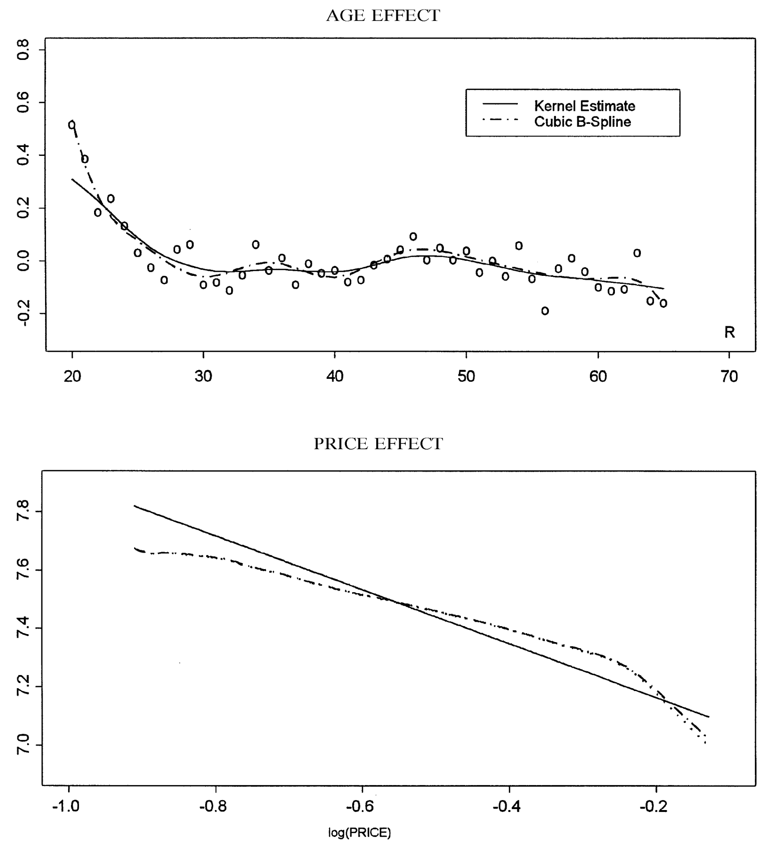


FIGURE 2.—Nonparametric price—partial linear model.[a]

fication tracks the kernel estimate closely in the central portion of the range of price but diverges at the extremes. The estimated log-linear price elasticity is $-.896$ while the average nonparametric price elasticity varies from $-.870$ to $-.910$ depending on the method of estimation.[8]

Having performed these preliminary tests, we are now in a position to test our semi-parametric specifications against pure parametric ones. A test of the specification

$$y = \gamma_0 + \gamma_1 \log(\text{PRICE}) + \gamma_2 bs(\text{AGE}) + z\beta + \varepsilon \tag{3.4}$$

against (3.1) yields $(10 \cdot 6230)^{1/2}(.4984 - .4953)/.4953 = 1.56$, which—keeping in mind that the test is one-sided—is not quite significant at the 5% level. A test of

$$y = \gamma_0 + \gamma_1 \log(\text{PRICE}) + \delta \text{AGE DUMMIES} + z\beta + \varepsilon \tag{3.5}$$

against (3.2) yields $(10 \cdot 6230)^{1/2}(.4964 - .4943)/.4943 = 1.06$.

In order to bolster confidence in the validity of our conclusions we repeated our estimation and inference procedures using various orders of differencing. Parameter estimates changed little and tests of significance and specification were consistent with the conclusions above. Application of Robinson's (1988) double residual method also led to very similar results. (These are available from the authors upon request.)

In addition, we implemented the conditional moment specification test proposed by Li (1994), Fan and Li (1996), Zheng (1996), and the goodness of fit statistic of Aït-Sahalia, Bickel, and Stoker (1998). These procedures resulted in similar conclusions. A test of joint significance of the two nonparametric variables (PRICE and AGE) resulted in a conditional moment statistic of 8.72 and a goodness of fit statistic of 10.8. Both are one-sided tests distributed as a standard normal under the null. A test of the pure parametric specification (3.4) against (3.1) yielded -1.36 and -1.24 for the two statistics. A test of (3.5) against (3.2) yielded values of -1.1 and -1.3. As our base case, we used band-widths of .2 for both log(PRICE) and log(AGE) when testing against specification (3.1). When testing against (3.2), we used a bandwidth of .05 for log(PRICE). Throughout we used a (product) normal kernel. Varying the bandwidth did not alter the overall conclusions.

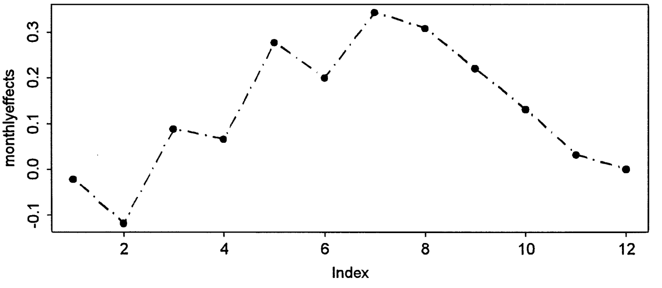
Finally, it is worth noting that across these various parametric and semi-parametric specifications that we have considered, there is relatively little variation in explanatory power—R^2 ranges from 26.6% to 27.2%.

3.3. *Demographic, Location and Temporal Effects*

Our estimated income elasticity is about .29, which lies midway between the value .37 estimated by Hausman and Newey (1995) and .21 obtained by Schmalensee and Stoker (1999) for U.S. data. Hausman and Newey's higher number is probably due to the fact that they are not conditioning on demographic effects. (Indeed, when we exclude the latter, our income elasticity also approximately doubles.) The number of drivers in the household is very strongly significant and the size of the household has a material and statistically significant effect. There is little change in these estimated effects across all the specifications considered in Figures 1 and 2.

The only important location variable is the urban dummy—families residing in cities drive about 33% less than their rural counterparts. Temporal effects are also significant

[8] Stoker (1992, Lecture 3) provides an excellent summary of direct and indirect average derivative estimators.



Monthly effects from cubic spline model in Figure 2.

FIGURE 3.—Estimated monthly effects.

as well as being consistent with intuition. The monthly dummies are plotted in Figure 3. Peak driving months occur during July and August when families drive 30% more than during December or January. The estimated May effect is marginally larger than the June effect probably because of the long weekend during May. In the U.S. the overall short-term peak for travel occurs during Thanksgiving in November. In Canada, Thanksgiving is a less celebrated holiday and occurs in October.

3.4. *Separate Estimation By Grade of Gasoline*

One of the interesting features of the Hausman and Newey estimates of the price effect is the variable curvature—price increases reduce demand at high and low prices but have a lesser impact at intermediate prices. Schmalensee and Stoker surmise that this is due to differences in the grade of gasoline and suggest that modeling "different types of gasoline might help explain differential price reactions." They do indeed find that the distribution of prices is bimodal (Schmalensee and Stoker (1999, p. 657–658)).

When analyzing data on *all* grades of gasoline collectively, we find no evidence of multi-modality in the distribution of prices (Figure 4, bottom panel, solid line) or the kind of variable curvature found by Hausman and Newey (Figure 2, bottom panel). However, when we estimate separate models for regular, medium, and premium gasoline[9] there is some evidence of 'flat zones' in the price effect (Figure 4, upper panel). Premium gas in particular exhibits a pattern similar to that found by Hausman and Newey.

Depending on the method of estimation, average nonparametric price elasticities vary from $-.878$ to -1.010 for premium gas, $-.369$ to $-.449$ for medium grade, and -1.088 to -1.155 for regular gas. Average elasticities for premium and regular grades thus appear to be similar even if the exact patterns differ (Figure 4, upper panel). Medium grade elasticities are much lower, but also less precisely estimated. A test of equality of average elasticities across the three grades was not rejected at conventional levels. However given

[9] These constitute 74%, 6%, and 20% of the sample respectively. Summary statistics are in Table II. It is not at all obvious that at this level of approximation one would want to estimate separate equations. (Based on a survey that we conducted, a relatively small proportion of car models *require* premium gasoline.) Those who do not require but still purchase premium gas probably do so because of perceived benefits in part resulting from advertising. Furthermore, endogeneity may be introduced when estimating separate equations since certain purchasers might alternate between grades depending upon price.

TABLE II

SUMMARY STATISTICS FOR PRICE DATA BY GRADE ($CAD/LITER)

Provinces	All Grades			Regular			Medium			Premium		
	mean	sd	n	mean	sd	n	mean	sd	n	mean	sd	n
Atlantic Provinces	0.605	0.054	1727	0.593	0.051	1251	0.610	0.046	123	0.646	0.046	353
Quebec	0.603	0.049	1231	0.588	0.039	963	0.630	0.035	89	0.667	0.045	179
Ontario	0.559	0.049	1142	0.547	0.039	943	0.582	0.047	57	0.628	0.045	142
Prairie Provinces	0.551	0.049	1638	0.547	0.046	1498	0.566	0.053	51	0.609	0.053	89
British Columbia	0.582	0.056	492	0.558	0.039	344	0.608	0.050	50	0.650	0.048	98
All Data	0.580	0.056	6230	0.567	0.049	4999	0.604	0.050	370	0.644	0.049	861

the large standard errors and relatively smaller samples for medium and premium grades, larger data sets could alter this conclusion. We also re-estimated our models allowing for different price elasticities at different income levels and found no evidence that price and income interact.

We now turn to estimates of parametric effects. Given the degree of precision of the estimates, equations for regular and premium gasoline yield similar results (Figure 4). Estimates for medium grade gasoline are somewhat different but these are based on a much smaller sample. A test of equality of parametric effects across all three grades is not rejected at either the 1% or 5% levels.

4. CONCLUSIONS

In this paper we estimate household demand for gasoline using Canadian micro-data. Our estimates of demographic effects are close to those found by Schmalensee and Stoker and our price elasticity, which is approximately $-.9$, is close to that found by Hausman and Newey. Our income elasticity is about .29, which is midway between those found by our predecessors. We did not find significant differences across equations when separate models were estimated for regular, medium, and premium grades of gasoline.

However, the *interpretation* of the price effect is in question. If one examines the variation in prices *within a given urban area* one finds that the coefficient of variation may be as much as 5% or higher for regular gasoline.[10] Thus, individuals who drive more are likely to encounter a broader range of prices and hence their search and transaction costs for cheap gasoline are lower. Furthermore, these same individuals derive greater benefit from cheap gas and therefore would be willing to incur higher search and transactions costs. Thus one might expect price to be negatively correlated with the residual in an equation where the dependent variable is distance traveled or the level of gasoline consumption. In this case the price coefficient would overestimate the true responsiveness of consumption to price.[11] To separate these two effects one should ideally have much more precise data on location. One could then instrument the observed price variable with the average price over a relatively small geographic area (such as the average intra-city price). This level of

[10] The coefficient of variation of the price of regular gasoline is about 9% in our complete data set. After adjusting for geographic and time of year effects, the coefficient falls to about 7%.

[11] In the extreme case, demand *could* be perfectly inelastic at the same time that the estimated price effect is significantly negative.

Partial Linear Model
Nonparametric Price[b]

	Regular Gas		Medium Gas		Premium Gas	
	coef	se	coef	se	coef	se
log income	0.2693	0.0240	0.3680	0.0795	0.2877	0.0537
log drivers	0.5393	0.0386	0.2955	0.1339	0.6240	0.0934
log hhsize	0.1131	0.0320	0.2590	0.1049	0.1001	0.0754
single, age < 35	0.1676	0.0703	0.2866	0.2292	0.4058	0.1708
Urban dummy	−0.3150	0.0227	−0.5648	0.0828	−0.3311	0.0542
average price elasticity[c]	−0.944	see below	−0.409	see below	−1.122	see below
R^2	.2623		.4331		.2886	
s^2	.5021		.3907		.4663	
Var(log(DIST))	.6806		.6772		.6555	
N	4999		370		861	

[a] Dependent variable is log(DIST). Order of differencing $m = 10$.
[b] Age effects estimated using cubic B-spline with knots at 20,25,30,35,40, 45,50,55,60,65. Monthly effects estimated using dummies.
[c] PRICE EFFECT: estimated average elasticities varied depending on method of estimation:

Premium: −.879 (se = .334) to −1.010 (se = .335);
Medium: −.369 (se = .388) to −.449 (se = .467);
Regular: −1.088 (se = .124) to −1.155 (se = .130);

[d] DISTRIBUTION OF LOG(PRICE): distribution for all grades, premium, medium, and regular grades estimated using kernel density estimator.

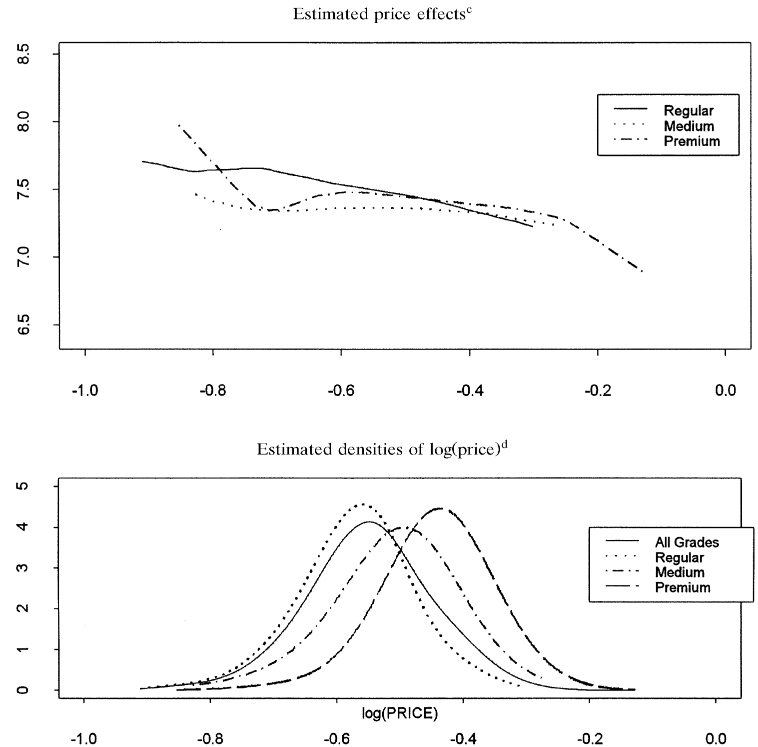


FIGURE 4.— Separate estimation by grade—partial linear model.[a]

detail was not available to us; however as a check on our estimates we instrumented our price variable with the five provincial/regional dummies. In the nonparametric version of the procedure we modified specification (3.2) to allow for a simple form of endogeneity as follows:

$$y = f(\text{PRICE}) + v\varphi + \delta \text{ AGE DUMMIES} + z\beta + \varepsilon \tag{4.1}$$

where v is defined by the instrumental variable equation $\text{PRICE} = \text{REGIONAL DUMMIES} \cdot \pi + v$ and $E(\varepsilon | \text{PRICE}, v, \text{AGE}, z) = 0$. (See Blundell and Duncan (1998) and Newey, Powell, and Vella (1999).) After estimating v from an OLS regression, equation (4.1) was estimated using differencing. The coefficient of v was .3454 with a standard error of .2488, which would not result in rejection of the null hypothesis that price is exogenous. Using instrumental variable estimation in the pure parametric specification (3.4) resulted in little change in the price coefficient and the Hausman (1978) test statistic ($\chi_1^2 = .53$) was also insignificant. Finally, it is worth pointing out that these tests may also be considered as tests of the presence of errors in measurement in the price variable. The low values of the statistics are consistent with the view that prices were measured with reasonable accuracy.

Department of Economics, University of Toronto, 150 St. George St., Toronto, Canada M5S 3G7; yatchew@chass.utoronto.ca

and

Department of Economics, University of Toronto, 150 St. George St., Toronto, Canada M5S 3G7; jyano@chass.utoronto.ca

Manuscript received November, 1999; final revision received August, 2000.

REFERENCES

AÏT-SAHALIA, Y., P. BICKEL, AND T. STOKER (1998): "Goodness-of-Fit Tests for Regression Using Kernel Methods," Manuscript, Princeton University.

BLUNDELL, R., AND A. DUNCAN (1998): "Kernel Regression in Empirical Microeconomics," *Journal of Human Resources,* 33, 62–87.

ESPEY, M. (1996): "Explaining the Variation in Elasticity Estimates of Gasoline Demand in the United States: A Meta-Analysis," *The Energy Journal,* 17, 49–60.

——— (1998): "Gasoline Demand Revisited: An International Meta-Analysis of Elasticities," *Energy Economics,* 20, 273–295.

FAN, Y., AND Q. LI (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica,* 64, 865–890.

HAUSMAN, J. (1978): "Specification Tests in Econometrics," *Econometrica,* 46, 1251–1271.

HAUSMAN, J., AND W. NEWEY (1995): "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss," *Econometrica*, 63, 1445–1476.

——— (1998): "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss," *Econometrics and Economic Theory in the 20th Century*. Cambridge: Cambridge University Press, 111–146.

LI, Q. (1994): "Some Simple Consistent Tests for a Parametric Regression Function versus Semiparametric or Nonparametric Alternatives," Department of Economics, University of Guelph, Manuscript.

NEWEY, W., J. POWELL, AND F. VELLA (1999): "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica,* 67, 565–603.

ROBINSON, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica,* 56, 931–954.

SCHMALENSEE, R., AND T. STOKER (1999): "Household Gasoline Demand in the United States," *Econometrica,* 67, 645–662.

STOKER, T. (1992): *Lectures on Semiparametric Statistics,* CORE, Louvain-la-Neuve (Belgium).

YATCHEW, A. J. (1997): "An Elementary Estimator of the Partial Linear Model," *Economics Letters,* 57, 135–143; additional examples contained in *Economics Letters* (1998), 59, 403–405.

——— (1998): "Nonparametric Regression Techniques in Economics," *Journal of Economic Literature*, 36, 669–721.

——— (1999): "Differencing Methods in Nonparametric Regression: Simple Techniques for the Applied Econometrician," Manuscript, University of Toronto.

ZHENG, J. (1996): "A Consistent Test of Functional Form via Nonparametric Estimation Techniques," *Journal of Econometrics*, 75, 263–289.