A Consistent Semiparametric Estimation of the Consumer Surplus Distribution[1]

Andrew Foster[2] Jinyong Hahn[3]

February, 1996

[1]We have benefitted from helpful comments by Joshua Angrist, Valentina Corradi, Christian Gourieroux, and the seminar participants of University of Pennsylvania, and Harvard/MIT. We are grateful to Jerry Hausman and Whitney Newey for kindly sharing the data.

[2]Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297

[3]Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6297

Abstract

In this paper, we examine the consequence of the random coefficient demand function in the estimation of the consumer surplus. It is shown that the distribution of the random coefficients can be consistently estimated when the demand function is linear in coefficients. Thus, the consumer surpluses distribution and the average consumer surplus can be consistently estimated even when the random coefficient distribution is not parametrically specified.

1 Introduction

Assume that the ith consumer's (Marshallian) demand is governed by the demand function $f_i(p, y)$: with price vector p and the income y, the consumer will spend $f_i(p, y)$. It is usually assumed that $f_i(\cdot)$ is the same across all consumers except for the allowance of intercept differences: $f_i(\cdot) = f(\cdot) + \epsilon_i$. Typically, a linear specification is used: $f_i(\cdot) = w_i'\theta + \epsilon_i$, where w_i denotes some vector-valued function of p and y. For this linear parametric specification, the consumer surplus is a function of θ, p^0, and p^1, say, $S(\theta, p^0, p^1)$. Determining $S(\cdot, \cdot, \cdot)$ involves solving a partial differential equation, but closed form expressions for some parametric specifications are available from Hausman (1981).

The above mentioned estimation strategy implicitly assumes that all consumers are homogeneous except for additive error terms. But this specification usually results in the rejection of the consumer rationality hypothesis. Brown and Walker (1989) argued that the rejection of the consumer rationality may be due to the heteroscedasticity of the disturbance term in the demand function. They argued that the random utility hypothesis may imply the disturbance term may not be homoscedastic, which would cause an inferential difficulties. With the interpretation of the disturbance term as the intercept heterogeneity of the individual demand functions, we are naturally led to consider the heterogeneous consumer preferences, where even the slope coefficients of the demand functions differ across individuals.

Heterogeneous consumer hypothesis necessitates a different way of thinking about the consumer surplus. Suppose that each consumer's (Marshallian) demand function is known up to some finite dimensional parameter, say $f(p, y, \theta)$. Also suppose that the consumer heterogeneity is solely summarized by θ. We can then view θ as a random vector, and consumer i's demand function *given* θ can be written as $f(\cdot, \cdot, \theta_i)$. Let $S(\theta, p^0, p^1)$ denote the consumer surplus change corresponding to the price change from p^0 to p^1 for the consumer with $\theta_i = \theta$. The price change from p^0 to p^1 would imply a consumer surplus change $S(\theta_i, p^0, p^1)$ for the i-th consumer. A social planner would in general like to know the complete distribution of $S(\theta_i, p^0, p^1)$ associated with the price change from p^0 to p^1, although some summary statistic suffices in many cases: a Benthamite utilitarian social planner, whose main concern is the total welfare, would want to know the average of the consumer surpluses

$E\left[S\left(\theta_i, p^0, p^1\right)\right]$.

If the variance of θ_i is strictly positive, estimation of the demand function under the assumption of consumer homogeneity will not in general yield the inference needed by the social planner. If (p, y) and θ are independent of each other in the population, then the econometrician would estimate $\mathbf{f}(p, y) \equiv E\left[f\left(p_i, y_i, \theta_i\right) \mid \left(p_i, y_i\right) = (p, y)\right]$, and then report the corresponding consumer surplus change. For example, if the demand function is linear in θ_i, then the econometrician would estimate $E[\theta_i]$, and report $S\left(E[\theta_i], p^0, p^1\right)$ as the consumer surplus change. It is easy to infer that $E\left[S\left(\theta_i, p^0, p^1\right)\right] \neq S\left(E[\theta_i], p^0, p^1\right)$, the answer that a utilitarian social planner would want.

As an illustration, consider a simple two-good economy consisting of two types of consumers. The Marshallian demand function for the first good of the first type of consumers is

$$q = 2 - \frac{1}{2}p + \frac{3}{2}y, \tag{1}$$

and that of the second type of consumers is

$$q = 4 - \frac{3}{2}p + \frac{1}{2}y, \tag{2}$$

where p is the price of the first good, and y is the income (taking the price of the second good as a numeraire). Suppose that the proportion of the first type in the economy is equal to $\frac{1}{2}$. Because the individual demand function satisfies the Gorman form, the average demand function is also linear

$$q = 3 - p + y. \tag{3}$$

Now, consider a price change from $p = 1$ to $p = 2$, and the corresponding equivalent variations for the two different types of individuals. For the general linear demand system

$$q = \alpha + \beta \cdot p + \gamma \cdot y,$$

the equivalent variation $EV\left(p^0, p^1, y; \alpha, \beta, \gamma\right)$ due to the price change from p^0 to p^1 equals

$$y - e^{\gamma \cdot \left(p^0 - p^1\right)} \cdot \left[y + \frac{1}{\gamma}\left(\beta \cdot p^1 + \frac{\beta}{\gamma} + \alpha\right)\right] + \frac{1}{\gamma}\left(\beta \cdot p^0 + \frac{\beta}{\gamma} + \alpha\right). \tag{4}$$

See Hausman (1981).[1] Thus, for the first type, the equivalent variation equals

$$\begin{aligned} & y - e^{-3/2} \cdot \left[y + \frac{2}{3} \cdot \left(-\frac{1}{2} \cdot 2 - \frac{1}{2} \cdot \frac{2}{3} + 2 \right) \right] + \frac{2}{3} \cdot \left(-\frac{1}{2} \cdot 1 - \frac{1}{2} \cdot \frac{2}{3} + 2 \right) \\ = & \left(1 - e^{-3/2} \right) \cdot y - e^{-3/2} \cdot \frac{4}{9} + \frac{7}{9}, \end{aligned}$$

and for the second type, the equivalent variation equals

$$\begin{aligned} & y - e^{-1/2} \cdot \left[y + 2 \cdot \left(-\frac{3}{2} \cdot 2 - \frac{3}{2} \cdot 2 + 4 \right) \right] + 2 \cdot \left(-\frac{3}{2} \cdot 1 - \frac{3}{2} \cdot 2 + 4 \right) \\ = & \left(1 - e^{-1/2} \right) \cdot y + e^{-1/2} \cdot 4 - 1. \end{aligned}$$

It thus follows that, under the assumption that the type is statistically independent of income, the average equivalent variation in the economy at a particular income level y equals

$$\left(1 - \frac{e^{-3/2} + e^{-1/2}}{2} \right) \cdot y - e^{-3/2} \cdot \frac{2}{9} + e^{-1/2} \cdot 2 - \frac{1}{9}. \tag{5}$$

Now, suppose that we only know the average demand function (3), and used (4) mechanically to compute the equivalent variation. We would then obtain

$$y - e^{-1} \cdot [y + (-2 - 1 + 3)] + (-1 - 1 + 3) = \left(1 - e^{-1} \right) \cdot y + 1, \tag{6}$$

which can be interpreted as the equivalent variation of the average consumer. Observe that (5) and (6) are clearly different from each other. A utilitarian social planner who posits a tax policy and thus wishes to know the average deadweight loss would prefer (5) to (6). Note, however, the tax revenue does not require any knowledge of the individual demand functions (1) and (2). If the marginal cost of production of the first good is constant at 1, the tax revenue can be computed from the average demand function (3) and equals $y + 1$. Thus, the difference between (5) and (6) solely reflects the difference in the deadweight loss calculation.

A concern about the difference between the expectation of the nonlinear function and the nonlinear function of the expectation is not new. For example, Brown and Mariano (1984) was concerned about this difference in the prediction context. Faced with the issue that the mean square error loss minimizing prediction does not have a closed form expression

[1] Hausman (1981) also calculated the compensating variation for this demand system.

for some nonlinear systems, they argued that a simulation based prediction yields the mean square loss minimizing prediction. They considered in particular two simulation methods. One method makes use of the parametric distributional assumption of the error term in the nonlinear system. They called it the Monte Carlo predictor. The other one is somewhat nonparametric in nature. It is based on the residual distribution: because the residual distribution is a consistent estimator of the error distribution, the resultant prediction is also consistent. They called it the residual-based predictor.

In this paper, we consider the estimation of the consumer surplus distribution when the demand function $f(p, y, \theta)$ is linear in θ: we assume that $f(p, y, \theta_i) = w'\theta_i$, where w is some function of (p, y)[2]. Assuming that the integrability condition is satisfied (or imposed), the identification of the consumer surplus distribution would be conceptually complete if the θ_i distribution were identified: with the θ_i distribution identified, $S(\theta_i, p^0, p^1)$ can be identified via simulation. The simulation would in general involve solving the partial differential equation for each realization of θ_i to obtain a corresponding $S(\theta_i, p^0, p^1)$. Furthermore, for some demand function specification, analytic solutions are available from Hausman (1981), which would substantially simplify the computation. In the linear demand example with

$$q_i = \alpha_i + \beta_i \cdot p + \gamma_i \cdot y,$$

the average equivalent variation for a price change from p^0 to p^1 for consumers with income y equals

$$\int EV\left(p^0, p^1, y; \alpha, \beta, \gamma\right) \, dF(\alpha, \beta, \gamma),$$

where F denotes the distribution of $(\alpha_i, \beta_i, \gamma_i)$. When F is parametrically specified, this average equivalent variation may be consistently estimated by parametric simulation. When $(\beta_i, \gamma_i) = (\beta, \gamma)$ is a constant, intercept variation summarizes all the consumer heterogeneity and (β, γ) can be consistently estimated. Thus, the α_i distribution can also be consistently estimated by the residual distribution, and the residual distribution from the regression along with the consistent slope coefficient estimators can be used to implement the residual based simulation. This methodology is *not* applicable, though, when the parameter distribution is

[2] We actually consider the case where some known transformation of the quantity demanded is linear in θ, e.g., the log linear demand system.

completely nonparametrically specified.

We consider the semiparametric estimation of the consumer surplus distribution and the corresponding average consumer surplus assuming that the θ_i distribution is nonparametrically specified. We show that the θ_i distribution can be identified and consistently estimated utilizing the methodologies of Beran and Hall (1992) and Beran and Millar (1994). For simplicity, we only consider the two-goods linear demand system:

$$q = \alpha_i + \beta_i \cdot p + \gamma_i \cdot y,$$

where, (q, p, y) can denote the logarithms of quantity demanded, price, and income. Observe that the closed form expression for the consumer surplus exists from Hausman (1981). Beran and Hall (1992) considered a simple linear regression model $Y_i = a_i + b_i X_i$, where a_i, b_i and X_i are independent of each other, and suggested a method for identifying the distributions of a_i and b_i. Beran and Millar (1994) generalized the result of Beran and Hall (1992). They considered a multiple linear regression model $Y_i = b_{1i} + \sum_{k=2}^{K} b_{ki} X_{ki}$, and suggested a method for estimating the joint distribution of $(b_{1i}, \ldots, b_{Ki})$ under the assumption that $(b_{1i}, \ldots, b_{Ki})$ is independent of $(X_{2i}, \ldots, X_{Ki})$. We will assume that the parameters $(\alpha_i, \beta_i, \gamma_i)$ are jointly independent of the price-income pair (p_i, y_i), and develop a method of identifying the joint distribution of $(\alpha_i, \beta_i, \gamma_i)$. For any demand function that is linear in the coefficients, the coefficient distribution can be similarly estimated. The distribution of the consumer surpluses is identified using the identified coefficient distribution.

We assume away the measurement error in this paper. In the linear demand function specification we work with, it is not clear how to separate out the measurement error of the quantity demanded and intercept heterogeneity, although Hausman (1985) shows that it is possible to separate them out in some nonlinear models. We also assume away the specification problem in this paper. We assume that the econometrician knows the functional form of the individual demand function except for the coefficient distribution. Ideally, we should consider some abstract "space" consisting of all the legitimate demand functions, and be able to identify the distribution of the demand functions on this space. Unfortunately, we were not able to develop a theory of this generality. Instead, we restrict our attention to the "space" consisting of the demand functions with some particular parametric specification, and consider the nonparametric identification of the distribution on this space. This approach is more restrictive in spirit than the one taken by Brown and Matzkin (1995), who consider some "space" of demand functions which are semiparametrically specified: they consider the utility functions consisting of a "nonparametric component" which is common across all the economic agents, and the parametric component with the parameter possibly different across economic agents. They assume ignorance about the nonparametric component and the distribution of the parameters. They then examine the nonparametric identification of the "nonparametric component" and the distribution of the parameters. Even though their approach is less restrictive than ours in spirit, their "space" of demand functions does not contain our "space" as a proper subset. We thus regard our result as complementary to theirs. Our approach also differ from the one taken by Hausman and Newey (1995). They assume the homogeneity of the consumers except possibly for the intercept, but assume ignorance about the functional form of this identical demand function. They go on to estimate this unknown demand function nonparametrically . The resultant consumer surplus estimator is interpreted as corresponding to some particular consumer type if the error term is assumed to be the intercept heterogeneity. The consumer surplus estimator is obviously the common consumer surplus for every consumer if the error term is assumed to be the measurement error. In that their "space" does not contain ours as a proper subset, we again regard our result as complementary to theirs.

This paper is organized as follows. In Section 2, we discuss the identification of the coefficient distribution based on the moments. This generalizes the result of Beran and Hall (1992). We then argue that the estimation of the coefficient distribution based on the estimated moments is not often practical. In Section 3, we discuss the estimation strategy of Beran and Millar (1994). Even though the estimation of Beran and Millar (1994) is more practical than the one based on the estimated moments as developed in Section 2, the identification based on moments is conceptually much easier to understand. We discuss both methods because of this trade-off although we use Beran and Millar's (1994) methodology in our estimation. In Section 4, we apply the technique in Section 3 to the gasoline data used in Hausman and Newey (1995). Section 5 summarizes the results.

2 Moment Based Identification

In this section, a consistent estimator of consumer surplus distribution is suggested for the two-good linear demand system. Even though we only discuss the linear demand system, the generalization to any demand system that is linear in coefficients is immediate. The estimator utilizes Hausman's (1981) calculation and some generalization of Beran and Hall's (1992) nonparametric estimation of the coefficient distribution in the random coefficient regression model.

Let p_i y_i denote the price of the first good and the income taking the price of the second good as the numeraire, respectively. Assume that the consumer i's demand of the first good is related to (p_i, y_i) through

$$q_i = \alpha_i + \beta_i p_i + \gamma_i y_i. \tag{7}$$

Consumer heterogeneity is summarized in the distribution of $\theta_i = (\alpha_i, \beta_i, \gamma_i)$, and we assume that the θ_i distribution is unknown to the econometrician. For the consumer with $\theta_i = \theta$ and $y_i = y$, the equivalent variation $EV\left(p^0, p^1, y; \theta\right)$ due to the price change from p^0 to p^1 is shown by Hausman (1981) to be

$$y - \exp\left[\gamma \cdot \left(p^0 - p^1\right)\right] \cdot \left[y + \frac{1}{\gamma}\left(\beta \cdot p^1 + \frac{\beta}{\gamma} + \alpha\right)\right] + \frac{1}{\gamma}\left(\beta \cdot p^0 + \frac{\beta}{\gamma} + \alpha\right). \tag{8}$$

Assume that the joint distribution of (θ_i, y_i) is identified. Because the marginal distribution of y_i is identified as the limit of the empirical distribution of y_i, this knowledge of the joint distribution amounts to the knowledge of the conditional distribution $F\left(\cdot\,|\, y_i\right)$ of θ_i given y_i. The distribution of the equivalent variation $EV\left(p^0, p^1, y; \theta\right)$ is easily seen to be the distribution of (8) under the product of $F\left(\cdot\,|\,\cdot\right)$ and the marginal distribution of y_i. The average compensating variation $E\left[CV\left(p^0, p^1, y_i, \theta_i\right)\right]$ is similarly identified. Because the marginal distribution of y_i is identified as the limit of the empirical distribution, it follows that the identification of $F\left(\cdot\,|\, y_i\right)$ suffices for the identification of the equivalent variation distribution.

We will assume that θ_i is independent of (p_i, y_i), and use the implied conditional moments of q_i to identify the θ_i distribution. This assumption is similar in spirit to Brown and Matzkin (1995), in which they assumed that observed characteristics are statistically independent of the unobserved ones. Notice that we can write $F(\cdot|\, y_i) = F(\cdot)$ because the independence assumption.

Assumption 1 (θ_i, p_i, y_i) $i = 1, 2, \ldots$ *is an i.i.d. sequence of random vectors defined on a probability space* $(\Omega, \mathcal{B}, P)$ *such that*

(i) θ_i *are independent of* (p_i, y_i).

(ii) The distribution of θ_i *is determined by its joint moments.*

(iii) The distribution of (p_i, y_i) *is characterized by a density which is bounded away from zero on some open set* O.

Remark: Assumption $(1.i)$ amounts to the assumption that the consumer preference summarized by θ_i is independent of (p_i, y_i). While the assumption θ_i is independent of p_i may be justified by assuming that the individual consumers are price takers, it is awkward to assume that θ_i is independent of y_i. After all, the income is *chosen* by the consumer to maximize the utility over consumption and leisure so that it is reasonable to expect some nonzero correlation between θ_i and y_i. Even if the income is defined as the wage rate times the total amount of time available and the leisure is regarded as a component of the consumption bundle, it is reasonable to believe that the consumer preference and the income are both correlated with some attributes x_i of the consumer. But if it is assumed that the coefficients are independently distributed of (p_i, y_i) *given* x_i, which may or may not be reasonable, the conditional distribution of the compensating variation given x_i and the conditional average compensating variation $E\left[EV\left(p^0, p^1, y_i, \theta_i\right)|\, x_i\right]$ can be estimated using the estimator of the conditional distribution of θ_i given x_i. Without the independence or the conditional independence between θ_i and (p_i, y_i), the estimator developed in this paper cannot be applicable for the consumer surplus distribution estimation. Assumptions $(1.ii)$ and $(1.iii)$ are crucial technical assumptions for identifying the distribution of θ_i.□

Rewrite the demand function as

$$q_i = \alpha + \beta \cdot p_i + \gamma \cdot y_i + (\alpha_i - \alpha_0) + p_i \cdot (\beta_i - \beta_0) + y_i \cdot (\gamma_i - \gamma_0),$$

where $\alpha_0 \equiv E[\alpha_i]$, $\beta_0 \equiv E[\beta_i]$, $\gamma_0 \equiv E[\gamma_i]$. Because of the independence between the regressor and the coefficients, the conditional expectation of

$$e_i \equiv (\alpha_i - \alpha_0) + p_i \cdot (\beta_i - \beta_0) + y_i \cdot (\gamma_i - \gamma_0)$$

given the regressor is equal to 0, so that $\theta_0 \equiv (\alpha_0, \beta_0, \gamma_0)$ can be consistently estimated by OLS. As observed by Beran and Hall (1992), the independence between the coefficients and the regressors imply a particular moment structure on the error term e_i. Because α_0, β_0, and γ_0 can be consistently estimated, we may assume without loss of generality that e_i distribution given (p_i, y_i) is identified. Define

$$\psi(t|p,y) \equiv E\left[e^{\sqrt{-1} t e_i} \middle| (p_i, y_i) = (p, y) \right].$$

Let

$$\psi^m(t) \equiv \frac{\partial^m \psi(t|p,y)}{\partial t^m}.$$

Letting $t \to 0$, we can identify

$$E\left[(u_i + p v_i + y w_i)^m\right] = \sum_{k_1+k_2+k_3=m} \frac{m!}{k_1! k_2! k_3!} p^{k_2} y^{k_3} E\left[u_i^{k_1} v_i^{k_2} w_i^{k_3}\right],$$

where

$$u_i = \alpha_i - \alpha_0, \quad v_i = \beta_i - \beta_0, \quad w_i = \gamma_i - \gamma_0.$$

For $m(m+1)(m+2)$ different pairs of (p, y), which we can choose because (p_i, y_i) has a positive density on O, we can obtain a system of $m(m+1)(m+2)$ linear equations with $m(m+1)(m+2)$ unknowns, $E\left[u_i^{k_1} v_i^{k_2} w_i^{k_3}\right]$ $k_1 + k_2 + k_3 = m$. Thus, all the joint moments of (u_i, v_i, w_i) are identified, and the distribution of (u_i, v_i, w_i) is identified because it is determined by its moments.

Theorem 1 *Under Assumption 1, the distribution of θ_i can be identified.*

Because the distribution F of θ_i is determined by its joint moments, we can estimate F via some consistent estimators of the joint moments

$$\mu(r, s, t) \equiv E\left[u_i^r v_i^s w_i^t\right], \quad r, s, t = 0, 1, 2, \ldots.$$

Observe that

$$\begin{aligned} E\left[e_i^k \middle| p_i, y_i\right] &= \sum_{r+s+t=k} \frac{k!}{r!s!t!} E\left[u_i^r p_i^s v_i^s y_i^t w_i^t \middle| p_i, y_i\right] \\ &= \sum_{r+s+t=k} \frac{k!}{r!s!t!} p_i^s y_i^t \mu(r,s,t). \end{aligned}$$

Were e_i observed, we can consistently estimate the joint moments $\mu(r,s,t)$ by regressing e_i^k on a constant and $p_i^s y_i^t$, $1 \leq s+t \leq k$. Because e_i is unobserved, we use instead $\hat{e}_i$, the OLS residual.

Assumption 2 *The covariance matrix of* $\kappa_k \equiv \operatorname{vec}\left[p_i^s y_i^t\right]_{1\leq s+t\leq k}$ *is nonsingular for every* $k = 2, 3, \ldots$.

The linear regression yields a consistent estimator $\hat{\mu}(r,s,t)$ of $\mu(r,s,t)$, $r+s+t \leq k$. Notice that $\hat{\mu}(r,s,t)$ is a strongly consistent estimator of $\mu(r,s,t)$ for each r,s,t. Let $\Delta_{r,s,t}$ denote the subset of Ω on which the convergence holds, and observe that $P[\Delta_{r,s,t}] = 1$. The convergence holds for every r,s,t on $\Delta \equiv \cap_{r,s,t}\Delta_{r,s,t}$, and $P[\Delta] = 1$. It follows that

Theorem 2 *Under Assumptions 1 and 2, there exists a subset* Δ *of* Ω *with* $P[\Delta] = 1$ *such that* $\hat{\mu}(r,s,t) \to \mu(r,s,t)$ a.s. *for every* r,s,t.

Now, let $\hat{F}_n$ denote *any* sequence of distribution functions whose r,s,t joint centered moments are equal to $\hat{\mu}(r,s,t)$ for $r+s+t \leq k(n)$, where $k(n)$ is some function of the sample size such that $\lim k(n) = \infty$.

Proposition 3 *Suppose that the distribution of* $U = \left(U^1, \ldots, U^K\right)$ *is determined by its joint moments, that the* $U_n = \left(U_n^1, \ldots, U_n^K\right)$ *have moments of all orders, and that*

$$\lim_{n\to\infty} E\left[\left(U_n^1\right)^{r_1} \cdots \left(U_n^K\right)^{r_K}\right] = E\left[\left(U^1\right)^{r_1} \cdots \left(U^K\right)^{r_K}\right] \qquad \forall r_1, \ldots, r_K.$$

Then, the distribution of U_n *converges weakly to that of* U.

Proof: The proof is a multivariate extension of Billingsley (1979, Theorem 30.2). Let ν_n and ν denote the distributions of U_n and U, respectively. Because $E\left[|U_n|^2\right]$ converges, it is bounded. Chebyshev's Inequality then implies that the sequence $\{\nu_n\}$ is tight. Because

ν_n is defined on a Euclidean space, which is completely separable, it is relatively tight by Prohorov's Theorem. See Billingsley (1968, Theorem 6.2), for example. Thus, given a subsequence $\{n'\}$, there exists a further subsequence $\{n''\}$ such that $\nu_{n''}$ converges to some limit distribution, say $\tilde{\nu}$. Let V denote a random vector whose distribution is given by $\tilde{\nu}$. Because joint moment of all orders exist for $U_{n''}$, the sequence $\{U_{n''}\}$ is uniformly integrable. (e.g. Billingsley (1979, p. 291)) Thus, $\lim_{n''\to\infty} E\left[(U_{n''}^1)^{r_1}\cdots\left(U_{n''}^K\right)^{r_K}\right] = E\left[(V^1)^{r_1}\cdots\left(V^K\right)^{r_K}\right]$ for all $r_1,\ldots,r_K$. But by hypothesis, $\lim_{n''\to\infty} E\left[(U_{n''}^1)^{r_1}\cdots\left(U_{n''}^K\right)^{r_K}\right] = E\left[(U^1)^{r_1}\cdots\left(U^K\right)^{r_K}\right]$ and ν is determined by its moments. Thus, it follows that $\tilde{\nu} = \nu$. It follows that ν_n converges weakly to ν. □

By Theorem 2 and Proposition 3, it follows that

Theorem 4 *$\hat{F}_n$ converges weakly to F almost surely.*[3]

Now, let H and $\hat{H}_n$ denote the population distribution and empirical distribution of y_i. By the Glivenko-Cantelli Theorem, $\hat{H}_n$ converges weakly to H almost surely. Because the sets of the form $\{t : t \leq s\}$ form a convergence-determining class in a finite-dimensional Euclidean space[4], the convolution of $\hat{H}_n$ and $\hat{F}_n$ converges weakly to that of F and H. Now, notice that the compensating variation is a continuous function of (θ_i, y_i). The Continuous Mapping Theorem then implies that

Theorem 5 *The distribution of the consumer surplus $EV(p^0, p^1, y; \theta)$ under the convolution of $\hat{F}_n$ and $\hat{H}_n$ converges weakly to the population distribution of the consumer surpluses almost surely.*

Because the estimated consumer surplus distribution converges weakly to the population consumer surplus distribution almost surely, it also follows that

Corollary 6 *The ξ-quantile of the consumer surplus $EV(p^0, p^1, y; \theta)$ under the convolution of $\hat{F}_n$ and $\hat{H}_n$ converges almost surely to the ξ-quantile of the population consumer surplus distribution.*

[3]See Giné and Zinn (1990) for the formal definition and properties of the almost sure weak convergence.

[4]Billingsley (1968, p. 18)

Because the estimated consumer surplus distribution converges weakly to the population consumer surplus distribution almost surely, it is tempting to assert that

$$\widehat{E}\left[EV\left(p^0, p^1\right)\right] \equiv \int\int EV\left(p^0, p^1, y; \theta\right) d\hat{H}_n(y) d\hat{F}_n(\theta)$$

would converge to the average consumer surpluses almost surely. But because the weak convergence is defined in terms of the *bounded* and continuous functions[5], certain regularity condition is necessary to rule out some pathological cases. We impose the following regularity condition.

Assumption 3 *There exists a compact set* $\mathbf{C}$ *such that* $F(\mathbf{C}) = 1$.

Remark: Although the assumption that the coefficient distribution is concentrated on some compact set may be reasonable assumption to make, the assumption that we have some knowledge of this support to be used in the estimation of the distribution is aesthetically and logically annoying. But because the weak convergence does not imply the convergence of expectations of potentially unbounded functions, we make this unpleasant assumption.□

From Theorem 4 and Assumption 3, it easily follows that the average consumer surplus estimator using the estimated coefficient distribution is strongly consistent for the average consumer surpluses.

Theorem 7 *Under Assumptions 1-3, if* $\hat{F}_n$ *is chosen such that* $\hat{F}_n(\mathbf{C}) = 1$ $n = 1, 2, \ldots$, *we have*

$$\widehat{E}\left[EV\left(p^0, p^1\right)\right] \to E\left[EV\left(p^0, p^1, y_i, \theta_i\right)\right] \qquad a.s.$$

So far, we have implicitly assumed that there exists a distribution $\hat{F}_n$ whose moments match the estimated ones $\hat{\mu}(r, s, t)$ for $r + s + t \leq k(n)$. But we are not guaranteed to have such a distribution. In other words, we are not guaranteed that the estimated moments would solve the moment problem. To discuss the moment problem, notice that not every sequence of numbers can be a sequence of valid moments. For example, if the square of the "second moment" is bigger than the "fourth moment", this sequence cannot be a sequence of valid moments. When a sequence of numbers is a sequence of valid moments, the sequence

[5] Billingsley (1968, p. 7)

is said to solve the moment problem. For the univariate case, von Mises (1964) provides a necessary and sufficient condition for a sequence of numbers to solve the moment problem. For the multivariate case, Cramér and Wold (1936) provides a sufficient condition.

Even though the consistency of $\hat{\mu}(r,s,t)$ implies that there exists a distribution with the matching moments when the sample size is sufficiently large, for practical purpose, it would be desirable to obtain $\hat{\mu}(r,s,t)$ as a solution to the constrained least squares estimator. Because $\mu(r,s,t)$ satisfy the necessary inequalities and solve the moment problem, such a constrained least squares estimator would still be strongly consistent, and we would be guaranteed the existence of a distribution with the matching moments. For the demand system where α_i, β_i, and γ_i are assumed to be independent of each other, finding the distributions with matching moments does not seem insurmountable. When α_i, β_i, and γ_i are independent of each other, we have

$$E\left[e_i^k \middle| p_i, y_i\right] = \sum_{r+s+t=k} \frac{k!}{r!s!t!} p_i^s y_i^t \mu_u(r)\, \mu_v(s)\, \mu_w(t),$$

where

$$\mu_u(r) = E\left[u_i^r\right], \quad \mu_v(s) = E\left[v_i^s\right], \quad \mu_w(t) = E\left[w_i^t\right].$$

In the univariate case, von Mises (1964) characterizes the necessary and sufficient inequalities for the sequence of numbers to be a valid sequence of moments. He shows that if a sequence of m numbers satisfy some m inequalities, there exist a distribution whose first m moments match those m numbers. Furthermore, he shows that there exists exactly one multinomial distribution with matching moments. Imposing these inequalities in the estimation of $\mu_u(r)$, $\mu_v(s)$, and $\mu_w(t)$ would result in the sequences of estimated moments which solve the moment problems. We can then use the algorithm described in Devroye (1986, p 686), for example, to construct a multinomial distribution with matching moments.

The assumption that α_i, β_i, and γ_i are independent of each other seems hard to justify economically, though. It would thus be desirable to find a condition under which $\widehat{\mu}(r,s,t)$ $r+s+t \leq k(n)$ are valid joint moments. So far, we have not been successful in finding an operational necessary and sufficient condition to be satisfied by $\mu(r,s,t)$ for this general multidimensional case. The sufficient condition discussed in Cramér and Wold (1936) involves every $\widehat{\mu}(r,s,t)$ $r,s,t = 1,2,\ldots$ Clearly, this type of condition is hard to impose in the estimation of moments.

3 Minimum Distance Estimation

In this section, we discuss a more practical estimation of the coefficient distribution of the linear demand system[6]

$$q_i = \alpha_i + \beta_i \cdot p_i + \gamma_i \cdot y_i.$$

This estimation strategy is taken from Beran and Millar (1994).

We first introduce some notations. Let F and G denote the distributions of $\theta_i = (\alpha_i, \beta_i, \gamma_i)$ and (p_i, y_i), respectively. Let $\mathcal{L}(F, G)$ denote the joint distribution of (q_i, p_i, y_i) under F and G. Define F_n, G_n, and $\mathcal{L}(F_n, G_n)$ to be some sequences of distributions of θ_i, (p_i, y_i), and (q_i, p_i, y_i). Finally, let d denote any metric that metrizes weak convergence, e.g., L_2-norm on characteristic functions.

To understand the intuition for the minimum distance estimation, consider the following variation of Beran and Millar's (1994) Proposition 2.2.

Proposition 8 *Suppose that Assumptions 1 and 3 holds. If*

$$\lim_{n\to\infty} d(G_n, G) = 0, \quad \lim_{n\to\infty} d(\mathcal{L}(F_n, G_n), \mathcal{L}(F, G)) = 0,$$

and $F_n(\mathbf{C}) = 1$ $n = 1, 2, \ldots,$ *then*

$$\lim_{n\to\infty} d(F_n, F) = 0.$$

Proof: This proof is virtually the same as Beran and Millar's (1994) proof of their Proposition 2.2. Because $\{F_n\}$ is tight, it has a subsequence converging weakly, say F_*. Assume that the random vector $\theta_* = (\alpha_*, \beta_*, \gamma_*)$ has this limit distribution F_*. Observe that $F_*(\mathbf{C}) = 1$. Because G_n converges weakly to G, it follows that $\mathcal{L}(F_n, G_n)$ converges weakly to $\mathcal{L}(F_*, G)$. We thus have

$$\mathcal{L}(F_*, G) = \mathcal{L}(F, G).$$

[6] Even though we only discuss the linear demand system, the generalization to any demand system which is linear in coefficients is immediate.

In terms of characteristic functions, this implies that

$$\begin{aligned} & E\left[\exp\left(\sqrt{-1}s\cdot(\alpha_i+\beta_i p_i+\gamma_i y_i)+\sqrt{-1}t\cdot p_i+\sqrt{-1}u\cdot y_i\right)\right] \\ = \; & E\left[\exp\left(\sqrt{-1}s\cdot(\alpha_*+\beta_* p_i+\gamma_* y_i)+\sqrt{-1}t\cdot p_i+\sqrt{-1}u\cdot y_i\right)\right] \end{aligned}$$

for all (s,t,u). This in particular implies that

$$\begin{aligned} & E\left[\exp\left(\sqrt{-1}s\cdot\alpha_i+\sqrt{-1}ps\cdot\beta_i+\sqrt{-1}ys\cdot\gamma_i\right)\right] \\ = \; & E\left[\exp\left(\sqrt{-1}s\cdot\alpha_*+\sqrt{-1}ps\cdot\beta_*+\sqrt{-1}ys\cdot\gamma_*\right)\right] \end{aligned} \tag{9}$$

for every (p,y) in the support of G. Because the distribution of (β_i,γ_i) is concentrated in the compact set, the left and the right sides of (9) are both analytic as functions of (p,y). Because G has a density bounded away from zero on some open set, (9) holds for every (s,p,y). Thus, $F=F_*$. Because every converging subsequence of $\{F_n\}$ converges weakly to F, it follows that F_n converges weakly to F.□

To define the minimum distance estimator, we introduce some additional notations. Let $\widehat{\mathcal{L}}_n$ and $\widehat{G}_n$ denote the empirical distributions of (q_i,p_i,y_i) and (p_i,y_i), respectively. We define the minimum distance estimator $\widehat{F}_n$ of F to be any distribution satisfying

$$d\left(\mathcal{L}\left(\widehat{F}_n,\widehat{G}_n\right),\widehat{\mathcal{L}}_n\right)\le\inf_{F_n} d\left(\mathcal{L}\left(F_n,\widehat{G}_n\right),\widehat{\mathcal{L}}_n\right)+o_p(1).$$

This minimum distance estimator is shown to be a consistent estimator of F in the following variation of Beran and Millar's (1994) Proposition 2.3.

Proposition 9 *Under Assumptions 1 and 3, we have*

$$\lim_{n\to\infty} d\left(\widehat{F}_n,F\right)=0.$$

Proof: Because $\widehat{G}_n$ converges weakly to G in probability, it follows that $\mathcal{L}\left(F,\widehat{G}_n\right)$ converges weakly to $\mathcal{L}(F,G)$. We thus have

$$d\left(\mathcal{L}\left(F,\widehat{G}_n\right),\mathcal{L}(F,G)\right)=o_p(1). \tag{10}$$

By the law of large numbers, we have

$$d\left(\widehat{\mathcal{L}}_n,\mathcal{L}\right)=o_p(1). \tag{11}$$

By the triangle inequality applied to (10) and (11), we thus obtain

$$\inf_{F_n} d\left(\mathcal{L}\left(F_n, \widehat{G}_n\right), \widehat{\mathcal{L}}_n\right) \leq d\left(\mathcal{L}\left(F, \widehat{G}_n\right), \widehat{\mathcal{L}}_n\right) = o_p(1).$$

It follows from the definition of $\widehat{F}_n$ that

$$d\left(\mathcal{L}\left(\widehat{F}_n, \widehat{G}_n\right), \widehat{\mathcal{L}}_n\right) = o_p(1). \tag{12}$$

By the triangle inequality applied to (11) and (12), we obtain

$$d\left(\mathcal{L}\left(\widehat{F}_n, \widehat{G}_n\right), \mathcal{L}(F, G)\right) = o_p(1).$$

Using the previous proposition, we obtain the desired conclusion.□

Because any $\widehat{F}_n$ such that

$$d\left(\mathcal{L}\left(\widehat{F}_n, \widehat{G}_n\right), \widehat{\mathcal{L}}_n\right) \leq \inf_{F_n} d\left(\mathcal{L}\left(F_n, \widehat{G}_n\right), \widehat{\mathcal{L}}_n\right) + o_p(1)$$

consistently estimates the θ distribution F, it would be convenient to restrict $\widehat{F}_n$ to be a multinomial distribution. Because $\hat{H}_n$ is also a multinomial distribution, a multinomial $\hat{F}_n$ reduces the computational cost of evaluating the equivalent variation distribution substantially compared to the continuous $\hat{F}_n$. This is especially true when the closed form expression for the compensating variation does not exist: with a multinomial $\hat{F}_n$, there are only a finite number of partial differential equation to be solved. For this purpose, let m_n denote any sequence of positive integers such that $\lim_{n\to\infty} m_n = \infty$. Let $M(m_n)$ denote the set of all multinomial distributions in $\mathbf{C}$ with mass at each point being m_n^{-1}. Letting

$$\widetilde{F}_n = \underset{F_n \in C(m_n)}{\operatorname{argmin}}\, d\left(\mathcal{L}\left(\widehat{F}_n, \widehat{G}_n\right), \widehat{\mathcal{L}}_n\right),$$

we still have the consistency of this minimum distance estimator, which is summarized in the following theorem due to Beran and Millar (1994, Proposition 2.4).

Theorem 10 *Under Assumptions 1 and 3, we have*

$$d\left(\widetilde{F}_n, F\right) = o_p(1).$$

Remark: The previous theorem does not give any guidance on the choice of m_n. As long as it approaches ∞ as the sample sizes increases, the consistency of the minimum distance estimator is not affected by the rate at which m_n increases. We do not yet know which rate provides the "optimal" behavior of the equivalent variation estimator under any sense of optimality. But a multinomial θ distribution has some convenient economic interpretation. The number of support points of this multinomial distribution can be interpreted as the number of consumer "types". A multinomial distribution with three support points is interpreted as indicating that there are three different types of consumers, for example. Thus, the determination of m_n in the finite sample may be guided by both economic and statistical intuition.□

The estimation of the minimum distance estimator requires a choice of $d(\cdot,\cdot)$. For this purpose, we can use the L_2-norm on the characteristic functions, for example. If L_2-norm on the characteristic functions is used, we have

$$d(P_1, P_2) = \left(\int |\phi_1(t) - \phi_2(t)|^2 \, dQ(t)\right)^{1/2},$$

where $\phi_1(t)$ and $\phi_2(t)$ are characteristic functions of P_1 and P_2, and Q is some probability with the support equal to the whole Euclidean space. It is often impossible to obtain an analytic expression of $d(\cdot,\cdot)$ when Q has the whole Euclidean space as its support. But we may instead use

$$d_N(P_1, P_2) = \left(\int |\phi_1(t) - \phi_2(t)|^2 \, dQ_N(t)\right)^{1/2},$$

where Q_N is the empirical distribution of a random sample of size N from Q. As long as $N \to \infty$ as $n \to \infty$, the corresponding simulated minimum distance estimator is would still be consistent. The estimation of the simulated minimum distance estimator $\widetilde{F}_n$ in practice thus consists of four steps:

1. Let
$$\widehat{\phi}(s,t,u) = \frac{1}{n}\sum_{i=1}^{n} \exp\left(\sqrt{-1}s \cdot q_i + \sqrt{-1}t \cdot p_i + \sqrt{-1}u \cdot y_i\right)$$
denote the empirical characteristic function of (q_i, p_i, y_i).

2. Let (a_k, b_k, c_k) $k = 1, \ldots, m$ denote the candidate support points for $\widetilde{F}_n$. Its character-

istic function is then equal to

$$\widetilde{\varphi}(s,t,u)=\frac{1}{m}\sum_{k=1}^{m}\exp\left(\sqrt{-1}s\cdot a_k+\sqrt{-1}t\cdot b_k+\sqrt{-1}u\cdot c_k\right).$$

The characteristic function $\ddot{\varphi}(s,t,u)$of $\mathcal{L}\left(\widetilde{F}_n,\widehat{G}_n\right)$ is then equal to

$$\ddot{\varphi}(s,t,u)=\frac{1}{n}\sum_{i=1}^{n}\widetilde{\varphi}(s,s\cdot p_i,s\cdot y_i)\exp\left(\sqrt{-1}t\cdot p_i+\sqrt{-1}u\cdot y_i\right).$$

3. Let Q_N denote the empirical distribution of a random sample of size N from Q. Evaluate

$$d_N=\int\left|\widehat{\phi}(s,t,u)-\ddot{\varphi}(s,t,u)\right|^2 dQ_N(s,t,u).$$

4. Minimize d_N over (a_k,b_k,c_k) $k=1,\ldots,m$.

4 Estimation

To estimate the average consumer surplus for gasoline, we use the data used by Hausman and Newey (1995). The data set is from the U.S. Department of Energy, and contains the monthly gasoline consumptions q_i, the weighted averages of the gasoline price over a month p_i, the incomes y_i, and other personal characteristics x_i of 18,109 observations. This personal characteristics x_i consists of 20 time and region indicator variables. For more complete description of the data set, see Hausman and Newey (1995, p. 1459). We use the log linear individual demand specification

$$\log q_i=\alpha_i+\beta_i\log p_i+\gamma_i\log y_i.$$

We assume that the personal characteristics x_i can be used to predict $\theta_i=(\alpha_i,\beta_i,\gamma_i)$, but the prediction is not perfect: we assume that

$$\begin{aligned}\alpha_i &= x_i'\alpha+a_i,\\ \beta_i &= x_i'\beta+b_i,\\ \gamma_i &= x_i'\gamma+c_i,\end{aligned}$$

where (a_i, b_i, c_i) has mean equal to zero and is independent of x_i. The distribution of (a_i, b_i, c_i) is assumed to be unknown. If (α, β, γ) is known, we can rewrite the individual demand function as

$$\begin{aligned} \log q_i^* &\equiv \log q_i - x_i'\alpha - (x_i'\beta) \cdot \log p_i - (x_i'\gamma) \cdot \log y_i \\ &= a_i + b_i \log p_i + c_i \log y_i. \end{aligned}$$

For this specification, the identification and consistent estimation of the (a_i, b_i, c_i) distribution can be achieved by the strategies discussed in the previous sections. Now, observe that (α, β, γ) can be consistently estimated by the OLS regression of $\log q_i$ on $(x_i, x_i \log p_i, x_i \log y_i)$, a standard linear regression model with interaction terms. Call the estimated OLS regression coefficient $\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}\right)$. Letting

$$\log \hat{q}_i \equiv \log q_i - x_i'\hat{\alpha} - \left(x_i'\hat{\beta}\right) \log p_i - (x_i'\hat{\gamma}) \log y_i,$$

and using it instead of $\log q_i^*$ would not cause any trouble in the consistent estimation of the (a_i, b_i, c_i) distribution because of the consistency of $\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}\right)$.

For the evaluation of the simulated metric $d_N(\cdot, \cdot)$, we need to choose Q and the simulated random sample size N: we chose $Q = N(0, I_3)$, and $N = 1,000$. We also need to choose m, the number of support points of $\widetilde{F}_n$: we experimented with $m = 4, \ldots, 10$. These choices of m can be interpreted as modelling the number of types of consumers for this particular data set. We believe that modelling the number of unobserved consumer heterogeneity to be between 4 and 10 would result in a fairly rich consumer heterogeneity because multiplying $m = 4$ to 10 unobserved characteristics with potentially 2^{20} support points of x_i will result in a lot of different consumer types.

For each choice of m, our estimation of the simulated minimum distance estimator $\widetilde{F}_n$ of the (a_i, b_i, c_i) distribution can be thus summarized as consisting of the following six steps:

1. Regress $\log q_i$ on $(x_i, x_i \log p_i, x_i \log y_i)$, and obtain the OLS regression coefficient $\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}\right)$.
2. Evaluate $\log \hat{q}_i \equiv \log q_i - x_i'\hat{\alpha} - \left(x_i'\hat{\beta}\right) \log p_i - (x_i'\hat{\gamma}) \log y_i$.
3. Define $\widehat{\phi}(s, t, u) = \frac{1}{n} \sum_{i=1}^n \exp\left(\sqrt{-1}s \cdot \log \hat{q}_i + \sqrt{-1}t \cdot \log p_i + \sqrt{-1}u \cdot \log y_i\right)$

4. Let $\left(\tilde{a}_k, \tilde{b}_k, \tilde{c}_k\right)$ $k = 1, \ldots, m$ denote the candidate support points for $\widetilde{F}_n$. Define

$$\widetilde{\varphi}(s,t,u) = \frac{1}{m}\sum_{k=1}^{m} \exp\left(\sqrt{-1}s \cdot \tilde{a}_k + \sqrt{-1}t \cdot \tilde{b}_k + \sqrt{-1}u \cdot \tilde{c}_k\right)$$

and

$$\ddot{\varphi}(s,t,u) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{\varphi}(s, s \cdot p_i, s \cdot y_i) \exp\left(\sqrt{-1}t \cdot p_i + \sqrt{-1}u \cdot y_i\right).$$

5. Let Q_N denote the empirical distribution of a random sample of size $1,000$ from $N(0, I_3)$. Evaluate

$$d_N = \int \left|\widehat{\phi}(s,t,u) - \ddot{\varphi}(s,t,u)\right|^2 dQ_N(s,t,u).$$

6. Minimize d_N over $\left(\tilde{a}_k, \tilde{b}_k, \tilde{c}_k\right)$ $k = 1, \ldots, m$. We did not impose the mean zero restriction of the (a_i, b_i, c_i) distribution in this minimization, though.

With the estimated (a_i, b_i, c_i) distribution $\widetilde{F}_n$ and the empirical distribution $\hat{H}_n$of the income distribution, we evaluated the distribution of the equivalent variation. It is the distribution of

$$EV\left(p^0, p^1, y_i; \theta_i\right) = y_i - \left\{(1-\gamma_i)\left[\frac{e^{\alpha_i}}{1+\beta_i}\left(p^{0\ 1+\beta_i} - p^{1\ 1+\beta_i}\right)\right] + y^{1-\gamma_i}\right\}^{\frac{1}{1-\gamma_i}}$$

under $\widetilde{F}_n \times \hat{H}_n$. We compared this distribution with the one where the only heterogeneity is through the observed characteristics x_i. In the latter distribution, we implicitly assumed that (a_i, b_i, c_i) are measurement errors, and computed the empirical distribution of $\left(\hat{\alpha}_i, \hat{\beta}_i, \hat{\gamma}_i\right)$, where

$$\begin{aligned}
\hat{\alpha}_i &= x_i'\hat{\alpha} + \frac{1}{m}\sum_{k=1}^{m} \tilde{a}_k, \\
\hat{\beta}_i &= x_i'\hat{\beta} + \frac{1}{m}\sum_{k=1}^{m} \tilde{b}_k, \\
\hat{\gamma}_i &= x_i'\hat{\gamma} + \frac{1}{m}\sum_{k=1}^{m} \tilde{c}_k.
\end{aligned}$$

The equivalent variation distribution was then computed under the product of this empirical distribution and $\hat{H}_n$.[7] Comparison of two equivalent variation distributions under these two

[7] Although we could have defined

$$\hat{\alpha}_i = x_i'\hat{\alpha}, \quad \hat{\beta}_i = x_i'\hat{\beta}, \quad \hat{\gamma}_i = x_i'\hat{\gamma},$$

different assumptions would tell us the extent to which the unobserved heterogeneity is significant.

Our findings are summarized in Tables 1-7. In general, we found that there is a substantial difference between the average equivalent variations computed under two assumptions. The distribution of the unobserved heterogeneity is substantial enough to make this difference the average of the equivalent variation, a nonlinear function of the parameters. For the price change from \$1.00 to \$ 1.30, we find that most of the difference can be accrued to the difference in the tax revenue calculation, but not to the deadweight loss. One possible conjecture is that the difference in the deadweight loss is minimal because the deadweight loss is of second order whereas the tax revenue is of the first order. For the price change from \$ 1.00 to \$ 1.50, though, we find some disturbing phenomenon. The difference between these two specifications seems to arise not only from the tax revenue but also from the deadweight loss. The average deadweight loss under the unobserved heterogeneity is even lower than the corresponding figure for the price change to \$ 1.30! In fact, for some types of consumers, the deadweight losses were estimated to be negative![8] This is one unpleasant feature for which we do not have too good an explanation. One possible reason may be that the standard deviation of the deadweight loss is so big that a reliable average may be hard to obtain even with a large sample. This is vindicated from our estimates of the standard deviation of the deadweight loss. A related possibility is that our estimator does not have a fast rate of convergence: even though we have established the consistency of our procedure, we do not expect the estimators to be $\sqrt{n}$-consistent. Last, it may simply be the case that the log linear demand system is not a good specification for the individual demand.

One might want to compare our estimates with Hausman and Newey's (1995). Their equivalent variation estimates are between \$ 278.95 and \$ 302.75 for the price change to \$ 1.30, and \$ 438.01 and \$ 475.91 for the price change to \$ 1.50.[9] Their deadweight loss

we wanted to check the sensitivity of our estimator for not imposing the zero mean restriction in the computation of $\widetilde{F}_n$.

[8]We verified that the estimated Hicksian demand elasticity was of the wrong sign for those types of consumer.

[9]We cannot make a direct comparison of their numbers with ours. They have computed the equivalent variation at a median income level, whereas we computed the average of the equivalent variation over over

estimates are between \$ 29.19 and \$ 38.68, and between \$ 45.80 and \$ 51.05, respectively. These numbers roughly corresponds to our estimates calculated under the assumption that all heterogeneity is observed. On the other hand, our equivalent variation estimates (computed under the assumption that some heterogeneity is unobserved) are between \$ 317.46 and \$ 332.69 for the price change to \$ 1.30, and \$ 513.85 and \$ 534.83 for the price change to \$ 1.50. This difference seems to suggest that the unobserved heterogeneity may be more important than the demand function specification. Our deadweight loss estimates for the price change to \$ 1.30 are roughly comparable to their numbers, whereas the ones for the price change to \$ 1.50 are not. Again, we do not very well understand why the numbers are so much different for the latter price change. Focusing on the former price change, though, we observe that most of the difference between our numbers and Hausman and Newey's (1995) can be attributed to the tax revenue calculation. Our conjecture is that the difference can be explained by the fact that they relied on the nonparametric regression of the *log* of the quantity demanded on price and income, and then solved for the implied demand. It is easy to see that this procedure would yield an inconsistent estimator of the average tax revenue unless the consumers are homogeneous.

5 Summary

For the linear demand function with the random coefficients, the issue of the consumer surplus distribution necessitates the estimation of the random coefficient distribution. Generalizing Beran and Hall (1992) and Beran and Millar's (1994) idea to the case where the random coefficients are not necessarily independent among themselves, it was established that the coefficient distribution can be identified and consistently estimated without making a parametric distributional assumption. When the random coefficients are independent of the regressor given some attributes of the consumer, the distributions of the random coefficients can be consistently estimated, so that the estimate of the consumer surplus distribution and the corresponding average consumer surplus based on these estimated coefficient distributions are consistent. The rate of convergence of such an estimator is not established, though.

different types of consumers and over different incomes levels.

This estimation methodology was applied to the gasoline data used in Hausman and Newey (1995). For this data set, it was found that the unobserved heterogeneity was significant enough to make a difference in the average equivalent variation.

References

Beran, R., and P. Hall (1992), "Estimating Coefficient Distributions in Random Coefficient Regressions," *Annals of Statistics*, Vol. 20, pp. 1970 - 1984.

Beran., and P. Millar (1994), "Minimum Distance Estimation in Random Coefficient Regression Models," *Annals of Statistics*, Vol. 22, pp. 1976 - 1992.

Billingsley, P. (1968), *Convergence of Probability Measures*, New York: Wiley.

Billingsley, P. (1979), *Probability and Measure*, New York: Wiley.

Brown, D., and R. Matzkin (1995) "Estimation of Nonparametric Functions in Simultaneous Equations Models, with an Application to Consumer Demand," *unpublished manuscript.*

Brown, B., and R. Mariano (1984) "Residual-Based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System," *Econometrica*, Vol. 52, pp. 321 - 343.

Brown, B., and M. Walker (1989), "The Random Utility Hypothesis and Inference in Demand Systems," *Econometrica*, Vol. 57, pp. 815 - 829.

Cramér, H. and H. Wold (1936), "Some Theorems on Distribution Functions," *Journal of the London Mathematical Society*, Vol. 11, pp. 290 - 294.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, New York: Springer-Verlag.

Giné, , E. and J. Zinn (1990), "Bootstrapping General Empirical Measures," *Annals of Probability*, Vol. 18, pp. 851-869.

Hausman, J. (1981), "Exact Consumer's Surplus and Deadweight Loss," *American Economic Review*, Vol. 71, pp. 662 - 676.

Hausman, J. (1985), "The Econometrics of Nonlinear Budget Sets," *Econometrica*, Vol. 53, pp. 1255-1282.

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

Pollard, D. (1984), *Convergence of Stochastic Processes*, New York: Springer-Verlag.

von Mises, R. (1964), *A Mathematical Theory of Probability and Statistics,* New York: Academic Press.

Table 1: $m = 4$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	317.46	513.85	290.34	453.84
	σ	134.57	238.84	82.88	129.85
Tax Revenue	μ	294.07	505.82	262.12	388.42
	σ	160.89	360.40	74.52	110.42
Deadweight Loss	μ	23.39	8.03	28.22	65.42
	σ	91.39	199.75	8.36	19.43

Table 2: $m = 5$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	318.56	517.43	288.99	452.73
	σ	140.96	252.30	82.89	130.15
Tax Revenue	μ	297.44	513.10	261.94	389.85
	σ	167.76	383.48	74.83	111.37
Deadweight Loss	μ	21.09	4.33	27.05	62.87
	σ	89.51	209.88	8.06	18.78

Table 3: $m = 6$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	320.51	520.22	287.88	451.66
	σ	149.71	261.14	82.84	130.27
Tax Revenue	μ	299.03	515.06	261.62	390.52
	σ	169.48	397.80	74.98	111.93
Deadweight Loss	μ	21.47	5.16	26.26	61.14
	σ	91.22	227.92	7.86	18.34

Table 4: $m = 7$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	334.00	528.39	293.18	451.10
	σ	186.32	307.77	83.52	128.80
Tax Revenue	μ	302.77	478.56	257.20	369.34
	σ	191.12	334.13	72.97	104.79
Deadweight Loss	μ	31.23	49.83	35.98	81.76
	σ	63.67	120.97	10.55	24.01

Table 5: $m = 8$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	326.41	527.52	287.08	450.73
	σ	168.02	287.14	82.48	129.80
Tax Revenue	μ	303.57	514.30	261.22	390.48
	σ	180.88	416.68	74.75	111.74
Deadweight Loss	μ	22.84	13.22	25.85	60.25
	σ	88.66	234.40	7.73	18.06

Table 6: $m = 9$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	329.59	531.72	286.88	451.55
	σ	177.04	301.83	82.62	130.05
Tax Revenue	μ	306.24	514.85	261.17	390.62
	σ	188.17	432.50	74.92	112.04
Deadweight Loss	μ	23.35	16.87	25.71	59.93
	σ	86.68	240.89	7.70	18.01

Table 7: $m = 10$

		Unobserved Heterogeneity		Observed Heterogeneity	
		\$1.0 – \$1.3	\$1.0 – \$1.5	\$1.0 – \$1.3	\$1.0 – \$1.5
Equivalent Variation	μ	332.69	534.82	287.16	450.27
	σ	184.14	313.78	82.69	129.95
Tax Revenue	μ	308.27	510.98	260.69	388.68
	σ	194.59	434.93	74.77	111.47
Deadweight Loss	μ	24.41	23.86	26.47	61.59
	σ	83.26	231.42	7.92	18.48