Econometrica Supplementary Material

SUPPLEMENT TO "NONPARAMETRIC ESTIMATION OF NONADDITIVE HEDONIC MODELS": TECHNICAL APPENDIX

(*Econometrica*, Vol. 78, No. 5, September 2010, 1569–1591)

BY JAMES J. HECKMAN, ROSA L. MATZKIN, AND LARS NESHEIM[1]

THIS TECHNICAL APPENDIX presents supplementary material. Appendix B discusses identification of general utility functions that are not linear in income. Appendix C discusses equilibrium bunching in hedonic models. Appendix D provides details and derives asymptotic results for the estimators proposed in the main paper. Appendix E presents some limited Monte Carlo evidence on the performance of the estimators. Appendix F presents proofs of the theorems in Appendix D.

APPENDIX B: IDENTIFICATION OF GENERAL UTILITY FUNCTIONS

Let utility take the form $U^*(I, z, x, \varepsilon)$, where $I = P(z) + R$ and R is nonwage income. Assume that $U_z^* = \frac{\partial U^*}{\partial z} < 0$ and $U_I^* = \frac{\partial U^*}{\partial I} > 0$, and as in the main paper assume that each worker has a unique interior optimum. Then the FOC of the worker becomes

$$P_z(z) = \frac{-U_z^*(P(z) + R, z, x, \varepsilon)}{U_I^*(P(z) + R, z, x, \varepsilon)}.$$

Under conditions analogous to those for the case where $U_I^* \equiv 1$, there exists a function

$$z = s(x, R, \varepsilon)$$

satisfying the first and second order conditions. Under the further assumption that $U_{I\varepsilon}^* P_z + U_{z\varepsilon}^* > 0$, this function is monotonically increasing in ε. So we can recover it as we do in the case where $U_I^* \equiv 1$.

[1]This research was supported by NSF Grants SES-0241858 and BCS-0433990/BCS-0852261. Lars Nesheim also thanks the Leverhulme Trust and the U.K. Economic and Social Research Council (ESRC Grant RES-589-28-0001) for support through its funding of the Centre for Microdata Methods and Practice (http://www.cemmap.ac.uk). Rosa L. Matzkin's research has been supported also by NSF Grant SES-0551272/SES-0833058. We are grateful to the co-editor, Whitney Newey, and three referees for their extremely useful input. We also thank the participants at seminars at Northwestern University, Universidad de San Andres, Princeton University, Harvard/MIT, UCLA/USC, University of California at Berkeley, University of Minnesota, the Bureau of Labor Statistics, University of Chicago, Universidad Torcuato Di Tella, the 2002 Workshop on Characteristics Models: Theory and Applications (University of Copenhagen), the 2002 Workshop on Mathematical Economics (IMPA), and the 2004 Banff International Research Station (BIRS) workshop on "Mathematical Structures in Economic Theory and Econometrics" for their useful comments. We thank Myrna Wooders and Daniel McFadden for many stimulating conversations, Donald J. Brown for helpful comments, and Yong Hyeon Yang for excellent research assistance.

 DOI: 10.3982/ECTA6388

Now define

$$\widetilde{m}(I, z, x, \varepsilon) = \frac{-U_z^*(I, z, x, \varepsilon)}{U_I^*(I, z, x, \varepsilon)}.$$

As in the case where $U_I^* = 1$, we need a dimension reduction to identify m. Suppose the function $\widetilde{m}$ can be written as

$$m(I, q, \varepsilon),$$

where q is a known function of z and $x \in \widetilde{X}_1 \subset \mathbf{R}$. Then we can identify m on the set $Q = \{(t_1, t_2, t_3) \in \mathbf{R}^2 \times \widetilde{E} \mid \text{for some } (x, R) \in \widetilde{X}_1 \times \mathbf{R}, t_1 = R + P(s(x, R, \varepsilon)) \text{ and } t_2 = q(s(x, R, \varepsilon), x)\}$.

To see this, fix numbers $(t_1, t_2, t_3) \in Q$. Since s can be identified as argued in the main text, we can find R^* and x^* such that

$$q(s(x^*, R^*, t_3), x^*) = t_2$$

and

$$P(s(x^*, R^*, t_3)) + R^* = t_1.$$

Then

$$m(t_1, t_2, t_3) = P_z(s(x^*, R^*, t_3)),$$

which shows that $m(t_1, t_2, t_3)$ is identified.

In summary, the conditions for identifying the ratio of U_z^* to U_I^* are analogous to those required to identify marginal utility when $U_I^* = 1$.

APPENDIX C: BUNCHING

Bunching in hedonic equilibrium arises when a positive measure of sellers or buyers (workers and firms) "bunch" and choose the same job type $z \in \widetilde{Z}$. When there is bunching, the equilibrium distribution of z has mass points.

Given a point $z_b \in \operatorname{int}(\widetilde{Z})$, it is only possible that a positive measure of agents will bunch at z_b if the set

$$B(z_b) = \{(x, \varepsilon), (y, \eta) \mid \Gamma_z(z_b, y, \eta) = U_z(z_b, x, \varepsilon)\}$$

has positive measure. For example, if $(x_b, \varepsilon_b, y_b, \eta_b) \in B(z_b)$ and the distributions of (x, ε) and (y, η) have mass points at (x_b, ε_b) and (y_b, η_b), then it is possible to have bunching at z_b in equilibrium. Alternatively, when (x, ε) and (y, η) are continuously distributed, the set $B(z_b)$ cannot have positive measure if $\Gamma_{z\eta} > 0$ and $U_{z\varepsilon} < 0$. This single-crossing condition rules out bunching and implies that (z, x, y) are continuously distributed on their respective domains.

Other examples of bunching can be generated either on the boundary of $\widetilde{Z}$ when positive masses of agents are at corner solutions or on the interior of $\widetilde{Z}$ if we relax the single-crossing condition. The next two subsections give examples of each of these possibilities.

C.1. *Bunching on the Boundary*

Let $\widetilde{Z} = [0, 1]$ and let $\Pi_0 = V_0 = 0$ so that reservation profits and utilities are zero. Suppose that each firm chooses z to maximize $z^\alpha \eta - P(z)$, where $\alpha = 0.5$ and $F_\eta(\eta) = \eta$ for $\eta \in [0, 1]$, and suppose that each worker maximizes $P(z) - z^\varepsilon$, where $F_\varepsilon(\varepsilon) = \frac{\varepsilon - 0.5\alpha}{1.5\alpha - 0.5\alpha}$ for $\varepsilon \in [\frac{\alpha}{2}, \frac{3\alpha}{2}]$. The first and second order conditions (SOC) for the firm are

$$\text{(C.1a)} \quad \text{FOC} \quad \alpha z^{\alpha-1}\eta - P_z(z) = 0,$$

$$\text{(C.1b)} \quad \text{SOC} \quad \alpha(\alpha-1)z^{\alpha-2}\eta - P_{zz}(z) < 0,$$

which implies that for those firms at an interior optimum,

$$\text{(C.2)} \qquad \eta(z) = \frac{P_z(z)z^{1-\alpha}}{\alpha}.$$

The first and second order conditions for the workers are

$$\text{(C.3a)} \quad \text{FOC} \quad P_z(z) - \varepsilon z^{\varepsilon-1} = 0,$$

$$\text{(C.3b)} \quad \text{SOC} \quad P_{zz}(z) - \varepsilon(\varepsilon-1)z^{\varepsilon-2} < 0.$$

For any interior equilibrium, we cannot have $\varepsilon < \alpha$. To see this, from the second order condition for the firm we obtain, after substituting (C.1a) into (C.1b) and collecting terms,

$$(\alpha - 1) < \frac{zP_{zz}(z)}{P_z(z)}.$$

From the second order conditions for the workers we obtain

$$P_{zz}(z) < \varepsilon(\varepsilon-1)z^{\varepsilon-2},$$

which is the same as

$$zP_{zz}(z) < \varepsilon(\varepsilon-1)z^{\varepsilon-1}.$$

Using the rewritten first order condition (C.3a), we can substitute $P_z(z)$ for $\varepsilon z^{\varepsilon-1}$ to obtain

$$\frac{zP_{zz}(z)}{P_z(z)} < \varepsilon - 1.$$

Thus $\varepsilon > \alpha$ is required to produce an interior solution.

In equilibrium, exactly half of all workers and firms choose the corner solution $z = 0$. The rest sort positively on the heterogeneity parameters (η, ε) and locate at interior optima. Each of the most productive firms is at an interior optimum (i.e., each of those firms with $\eta > \alpha = \frac{1}{2}$) and each of the high elasticity workers (the ones with low disutility of effort) participates at an interior ($\varepsilon > \alpha = \frac{1}{2}$). Since we assume that $z \leq 1$, the high elasticity persons are the ones who have the least disutility of work.

Since there is positive assortative matching,

$$\eta(\varepsilon) = F_\eta^{-1}(F_\varepsilon(\varepsilon)), \quad \varepsilon \in \left[\alpha, \frac{3\alpha}{2}\right].$$

Using our specific functional forms for the distributions, we obtain

$$\eta(\varepsilon) = \frac{\varepsilon - \frac{\alpha}{2}}{\alpha} = \frac{\varepsilon}{\alpha} - \frac{1}{2}, \quad \varepsilon \in \left[\alpha, \frac{3\alpha}{2}\right]. \tag{C.4}$$

Then using first order conditions (C.1a) and (C.3a), we obtain

$$\varepsilon z^{\varepsilon - 1} = \alpha \eta z^{\alpha - 1}.$$

Substituting $\eta(\varepsilon)$ in this expression, we obtain

$$z = \left(1 - \frac{\alpha}{2\varepsilon}\right)^{1/(\varepsilon - \alpha)}, \quad \varepsilon \in \left[\alpha, \frac{3\alpha}{2}\right]. \tag{C.5}$$

This is the equilibrium demand function. The matching supply function can be calculated by using (C.4) to substitute out for ε in (C.5). As a consequence, the interval with positive density of demand and supply is $Z = [0, (\frac{2}{3})^4]$. No closed form solution for the price function exists, but we can characterize the marginal price function using (C.3a) and (C.5); in particular, as $\varepsilon \to \alpha$, $z \to 0$ and $P_z(z)$ becomes arbitrarily large. This is an equilibrium because the supply density equals the demand density at each interior z. Consumers and firms not at the boundary are each at an interior optimum in this interval.

C.2. *Bunching on the Interior*

The previous section gives conditions that produce bunching on the boundary of the space of feasible attributes. In equilibrium, a positive fraction of agents do not have an interior optimum. Bunching on the interior occurs when a positive fraction of both workers and firms have an optimum at a single point in the interior of $\widetilde{Z}$. To produce bunching at z_b, the set of workers who satisfy

$$P_z(z_b) - U_z(z_b, x, \varepsilon) = 0$$

and the set of firms that satisfy

$$\Gamma_z(z_b, y, \eta) - P_z(z_b) = 0$$

must both have positive measure. If U_z and Γ_z are differentiable and the distributions of (x, ε) and (y, η) are absolutely continuous with respect to Lebesgue measure, this can only happen at z_b if the set

$$B(z_b) = \{(y, \eta, x, \varepsilon) \mid \Gamma_z(z_b, y, \eta) = U_z(z_b, x, \varepsilon)\}$$

has dimension $n_x + n_y + 2$. The set of agents who choose z_b in equilibrium is a subset of $B(z_b)$. If $B(z_b)$ has dimension less than $n_x + n_y + 2$, then it has measure zero and the set of agents who choose z_b has measure zero.

An alternative way to see this is to note that if there is bunching at z_b in equilibrium, then

$$z_b = d(y, \eta) = s(x, \varepsilon)$$

for sets of (y, η) and (x, ε) of equal and positive measure. This means that

$$\frac{\partial d(y, \eta)}{\partial y} = \frac{\partial d(y, \eta)}{\partial \eta} = 0,$$
$$\frac{\partial s(x, \varepsilon)}{\partial x} = \frac{\partial s(x, \varepsilon)}{\partial \varepsilon} = 0$$

for $\{(y, \eta, x, \varepsilon) \mid z_b = d(y, \eta) = s(x, \varepsilon)\}$.

To see how interior bunching might arise, consider the following example. Let y measure managerial skill or quality and let the distribution of manager skill be given by the distribution function F_y such that y is a continuous random variable. Let z measure hours of work on a job. A manager of type y has a production function that is quadratic in z:

$$\Gamma = \left\{ \begin{array}{ll} \Gamma_0 + \Gamma_1(y)z + \Gamma_4(y)z^2, & y \in [y_0, y_1) \\ \Gamma_0 + \Gamma_2(y)z + \Gamma_4(y)z^2, & y \in [y_1, y_2) \\ \Gamma_0 + \Gamma_3(y)z + \Gamma_4(y)z^2, & y \in [y_2, y_3] \end{array} \right\},$$

where $\Gamma_4(y) < 1$ for all y,

$$\text{(C.6)} \quad \begin{aligned} \Gamma_1(y) &= (-25F_y(y)^2 + 10F_y(y) + 1)(1 - \Gamma_4(y)), \\ \Gamma_2(y) &= 2(1 - \Gamma_4(y)), \\ \Gamma_3(y) &= (25F_y(y)^2 - 40F_y(y) + 18)(1 - \Gamma_4(y)), \end{aligned}$$

and $y_0 = F_y^{-1}(0)$, $y_1 = F_y^{-1}(0.2)$, $y_2 = F_y^{-1}(0.8)$, and $y_3 = F_y^{-1}(1)$. Assuming that F_y is twice continuously differentiable, this production function is twice continuously differentiable in all arguments and is quadratic in z. Over the relevant range of z, managers of higher quality have higher marginal productivity.

On the worker side, let x measure disutility from work and let the distribution of worker types have distribution function F_x such that x is a continuous random variable. Suppose utility for a worker with characteristic x is

$$U = \left\{ \begin{array}{ll} U_0 + U_1(x)z + U_4(x)z^2, & x \in [x_0, x_1) \\ U_0 + U_2(x)z + U_4(x)z^2, & x \in [x_1, x_2) \\ U_0 + U_3(x)z + U_4(x)z^2, & x \in [x_2, x_3] \end{array} \right\},$$

where $1 < U_4(x)$ for all x,

$$\text{(C.7)} \qquad \begin{aligned} U_1(x) &= (-25F_x(x)^2 + 10F_x(x) + 1)(1 - U_4(x)), \\ U_2(x) &= 2(1 - U_4(x)), \\ U_3(x) &= (25F_x(x)^2 - 40F_x(x) + 18)(1 - U_4(x)), \end{aligned}$$

and $x_0 = F_x^{-1}(0)$, $x_1 = F_x^{-1}(0.2)$, $x_2 = F_x^{-1}(0.8)$, and $x_3 = F_x^{-1}(1)$. As with firms, this utility function is quadratic in z and twice continuously differentiable. Over the relevant range of z, workers with higher values of x have lower marginal disutility of work.

This example generalizes the seminal Tinbergen (1956) normal-quadratic hedonic model. The equilibrium price function in this economy can be shown to be

$$P(z) = p_0 + z^2.$$

When y is uniformly distributed so that $F_y(y) = y$ for $y \in [0, 1]$, the demand function is as shown in Figure 1. For y in the interval $[0.2, 0.8]$, the first order

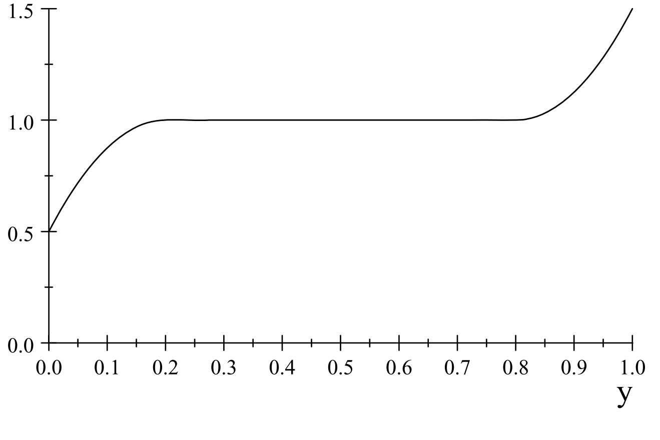

FIGURE 1.—Demand for hours of work.

condition for the firm is

$$2(1 - \Gamma_4(y)) + 2\Gamma_4(y)z = 2z$$

or

$$1 - \Gamma_4(y) = (1 - \Gamma_4(y))z,$$

so $z = 1$ is optimal for all y in this interval ($\frac{\partial d(y)}{\partial y} = 0$ in this interval). Similarly, for x in the interval $[0.2, 0.8]$, the first order condition for the worker is

$$2(1 - U_4(x)) + 2U_4(x)z = 2z$$

and again $z = 1$ is optimal for all x in the interval.

In this example, 60% of the managers and 60% of the workers choose to bunch at $z = 1$. Tangency conditions for two managers with particular values of y in the interval $[0.2, 0.8]$ are shown in Figure 2. Both indifference curves shown in the figure are tangent to the hedonic price at $z = 1$. The two indifference curves have different curvatures at this point. In fact, there is a full cluster of indifference curves with positive probability mass that are tangent to the price function at $z = 1$.

Over all intervals of y, the demand for z by firms of type y is

$$z(y) = \left\{ \begin{array}{ll} -\frac{25}{2}F_y(y)^2 + 5F_y(y) + \frac{1}{2}, & y \in [y_0, y_1) \\ 1, & y \in [y_1, y_2) \\ \frac{25}{2}F_y(y)^2 - 20F_y(y) + 9, & y \in [y_2, y_3] \end{array} \right\}.$$

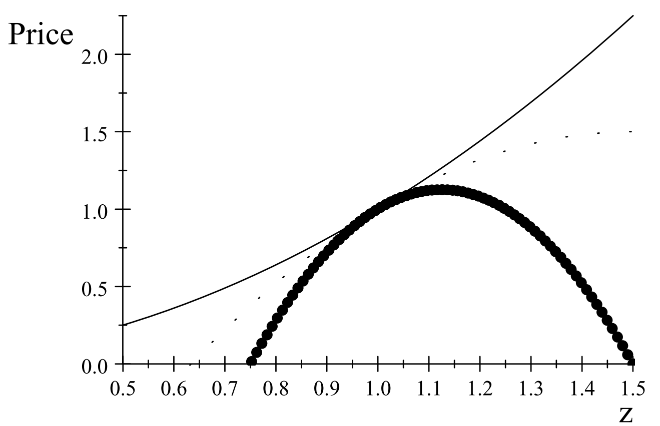

FIGURE 2.—Equilibrium bunching: tangency to hedonic price.

The supply function is similar. All managers with skill less than y_1 (20% of the population) employ part-time workers ($z < 1$). All managers with skill greater than y_1 and less than y_2 (60% of the population) employ full-time workers ($z = 1$), and all managers with skills greater than y_2 employ workers who work overtime ($z > 1$). Similarly, 60% of the workforce bunch at $z = 1$ or at full-time work. In this model, those choosing $z = 1$ are the mediocre managers and the mediocre workers.[2]

Such bunching is a knife-edge phenomenon. Any perturbation of the price function (so that the term in z is not quadratic or does not have a unitary coefficient) will break the bunching. So will choice of a more general coefficient on the linear term of the quadratic technologies.

APPENDIX D: ESTIMATION

We describe how to use the identification theorems from Section 3 in the main paper to develop consistent estimators for the underlying functions and distributions. We focus on models satisfying the conditions of Theorem 3.3 because of its generality. We assume that all the conditions made in the main paper are satisfied, and, for simplicity, that $\widetilde{X}_3$ is empty while $\widetilde{X}_1, \widetilde{X}_2 \subset R$. In Section D.2, we show how to alter the argument to define a consistent estimator for a model satisfying the conditions of Theorem 3.1 when $\widetilde{X}_2$ is empty while $\widetilde{X}_1 \subset R$. Estimators for models satisfying conditions in Theorem 3.2 can be defined analogously.

To focus on the properties of the estimator for the structural function and to keep the notation simple, we assume that $P_z(z)$ is known. In practice, $P_z(z)$ can be estimated by nonparametric kernel regression. If the dimension of x in the function m is strictly larger than that of $z + 2$, the rate of convergence of the estimator for $P(z)$ may be made faster than the rate for the structural function m. Hence, the variance due to the estimation of $P_z(z)$ would not affect the asymptotic distribution of the estimator for m. To simplify notation, we do not carry x_3 in our equations below.

[2]In this example, the bunching point $z = 1$ is determined for exogenous technological reasons. Such a bunching point could also emerge endogenously due to social coordination. For example, suppose that the production function were as above, but utility depended on $E(z)$, the average level of z in the market. In particular, in the previous example, replace (C.7) with

$$\begin{aligned}
U_1(x) &= \big(25[1 - 2E(z)]F_x(x)^2 - 10[1 - 2E(z)]F_x(x) + 1\big)(1 - U_4(x)),\\
U_2(x) &= 2E(z)(1 - U_4(x)),\\
U_3(x) &= \big(25[3 - 2E(z)]F_x(x)^2 - 40[3 - 2E(z)]F_x(x) + 6[8 - 5E(z)]\big)(1 - U_4(x)).
\end{aligned}$$

In this case, equilibrium bunching again emerges with 60% of the population choosing $z = E(z) = 1$. In this example, $E(z) = 1$, so in a single cross section, the model in the text and this model are indistinguishable.

Suppose that the available data are $\{Z^i, X^i\}$, $i = 1, \dots, N$, and that $P(z)$ is known. Let $f(z, x)$ and $F(z, x)$ denote, respectively, the joint probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of (Z, X). Let $\hat{f}(z, x)$ and $\hat{F}(z, x)$ denote the corresponding kernel estimators. Let $\hat{f}_{Z|X=x}(z)$ and $\hat{F}_{Z|X=x}(z)$ denote the kernel estimators of, respectively, the conditional p.d.f. and the conditional c.d.f. of Z given $X = x$. In this notation,

$$\hat{f}(z, x) = \frac{1}{N\sigma_N^{n_x+1}} \sum_{i=1}^{N} K\left(\frac{z - Z^i}{\sigma_N}, \frac{x_1 - X_1^i}{\sigma_N}, \dots, \frac{x_{n_x} - X_{n_x}^i}{\sigma_N}\right),$$

$$\hat{F}(z, x) = \int_{-\infty}^{z} \int_{-\infty}^{x} \hat{f}(s, t)\, ds\, dt,$$

$$\hat{f}_{Z|X=x}(z) = \frac{\hat{f}(z, x)}{\int_{-\infty}^{\infty} \hat{f}(s, x)\, ds},$$

$$\hat{F}_{Z|X=x}(z) = \frac{\int_{-\infty}^{z} \hat{f}(s, x)\, ds}{\int_{-\infty}^{\infty} \hat{f}(s, x)\, ds},$$

where $K : \mathbf{R}^{n_x+1} \to \mathbf{R}$ is a kernel function and σ_N is the bandwidth. For any t and x, $\hat{F}_{Z|X=x}^{-1}(t)$ will denote the set of values of Z for which $\hat{F}_{Z|X=x}(z) = t$. When the projection of the kernel function K onto Z is positive for all $z \in Z_s$, $\hat{F}(z, x)$ is uniquely determined for all z and each value of x.[3]

D.1. *Case 1*

Theorem 3.3 assumes that the marginal utility function is weakly separable into two functions that depend one on x_1 and the other on x_2; that is,

$$U_z(z, x_1, x_2, \varepsilon) = m(q_1(z, x_1), q_2(x_2, \varepsilon)),$$

where m is strictly increasing in its second argument, $q_1 : \mathbf{R}^2 \to \mathbf{R}$ and $q_2 : \mathbf{R}^2 \to \mathbf{R}$ are *known* functions, and q_2 is strictly decreasing in its second argument. Following the conditions of Theorem 3.3, normalize the value of the function m at one point $(\overline{z}, \overline{x}_1, \overline{\alpha})$ by requiring that

$$m(q_1(\overline{z}, \overline{x}_1), \overline{\alpha}) = P_z(\overline{z}).$$

[3] Z_s is the support of the mapping $s(x, \varepsilon)$ in equilibrium.

Define $s(x_1, x_2, \varepsilon)$ to be the solution to the workers' FOC. As noted in the proof of Theorem 3.3, from the assumed structure of separability, we can write

$$s(x_1, x_2, \varepsilon) = v(x_1, q_2(x_2, \varepsilon))$$

for some unknown function v that satisfies the property that

$$v(\overline{x}_1, \overline{\alpha}) = \overline{z}.$$

Since s is strictly increasing in ε, v is strictly decreasing in its second argument.

For any e, let $\overline{x}_2$ be such that $q_2(\overline{x}_2, e) = \overline{\alpha}$. From Matzkin (2003), it follows that

$$\widehat{F}_\varepsilon(e) = \widehat{F}_{Z|X=(\overline{x}_1,\overline{x}_2)}(\overline{z}).$$

Further, for any $\widetilde{x}_1, \widetilde{x}_2, \widetilde{e}$,

$$\widehat{v}(\widetilde{x}_1, q_2(\widetilde{x}_2, \widetilde{e})) = \widehat{F}^{-1}_{Z|X=(\widetilde{x}_1,\widetilde{x}_2)}(\widehat{F}_\varepsilon(\widetilde{e})).$$

This defines an estimator for v.

Let (t_1, t_2) be arbitrary points with $t_2 \in [q_2^l, q_2^u]$ and $t_1 \in \widetilde{Q}(t_2)$. To obtain an estimator for $m(t_1, t_2)$, first calculate $\widehat{x}_1^*$, the value that satisfies

$$q_1(\widehat{v}(\widehat{x}_1^*, t_2), \widehat{x}_1^*) = t_1.$$

Define $\widehat{m}(t_1, t_2)$ by

$$\widehat{m}(t_1, t_2) = P_z(\widehat{v}(\widehat{x}_1^*, t_2)). \tag{D.1}$$

Theorem D.1 establishes the asymptotic properties of this estimator for the case where the function $q_1(z, x_1) = zx_1$ and the function $q_2(x_2, \varepsilon) = x_2 - \varepsilon$. We offer this analysis as a prototype for a broad class of estimators. Similar results can be obtained for other specifications of the functions q_1 and q_2. Note that in this case, $n_x = 2$. Let $B(t, \xi)$ denote the neighborhood centered at t and with radius $\xi > 0$.

Given (t_1, t_2), let x_2^* and e^* be such that $x_2^* - e^* = t_2$ and $x_2^* \in \widetilde{X}_2$. Let x_1^* solve

$$q_1(v(x_1^*, t_2), x_1^*) = t_1$$

and let $\overline{x}_2 = \overline{\alpha} + e$. Define $x^* = (x_1^*, x_2^*)$. We assume that $(x_1^*, x_2^*) \neq (\overline{x}_1, \overline{x}_2)$.

We make the following assumptions:

ASSUMPTION A.1: The sequence $\{Z^i, X^i\}$, $i = 1, \dots, N$ is independent and identically distributed (i.i.d.).

ASSUMPTION A.2: $f(z, x)$ has compact support, which includes $\widetilde{Z} \times \widetilde{X} \subset \mathbf{R}^{1+n_x}$, and is continuously differentiable of order $s' \geq 4$.

ASSUMPTION A.3: The kernel function K is differentiable of order $\widetilde{s}$, the derivatives of K of order $\widetilde{s}$ are Lipschitz, and K vanishes outside a compact set, integrates to 1, and is of order s'', where $\widetilde{s}+s''+1\le s'$.

ASSUMPTION A.4: As $N\to\infty$, $\sigma_N\to 0$, $\ln(N)/N\sigma_N^{n_x+1}\to 0$, $\sqrt{N\sigma_N^{n_x}}\to\infty$, $\sqrt{N\sigma_N^{n_x+2s''}}\to 0$, and $\sqrt{N\sigma_N^{n_x}}(\sqrt{\ln(N)/(N\sigma_N^{n_x+7})}+\sigma_N^{s''})^2\to 0$.

ASSUMPTION A.5: $x_1^*\neq 0$; $0<f(x^*), f(\overline{x})<\infty$; there exist $\delta,\xi>0$ such that $\forall(x_1',x_2')\in B((x_1^*,x_2^*),\xi)$, $\forall\widetilde{z}\in B(v(x_1^*,t_2),\xi)$, $f(x_1',x_2')\ge\delta$ and $f(\widetilde{z},x_1',x_2')\ge\delta$; there exist $\delta',\xi'>0$ such that $\forall(x_1',x_2')\in B((\overline{x}_1,\overline{x}_2),\xi')$, $\forall\widetilde{z}\in B(v(\overline{x}_1,\overline{\alpha}),\xi')$, $f(x_1',x_2')\ge\delta'$ and $f(\widetilde{z},x_1',x_2')\ge\delta'$; and on all x in a neighborhood of x^*, $dF_{Z|X=x}(t_1/x_1)/dx_1$ is bounded away from 0.

ASSUMPTION A.6: $t_1\neq 0$ and it belongs to the interior of the support of $q_1(v(x_1,t_2),x_1)$.

Assumptions A.1, A.3, and A.4 are standard conditions on the data, the kernel, and the bandwidth when the estimators are functionals of kernel estimators for the distributions of the observed variables.[4] They imply no further economic restrictions on the model. In contrast, Assumptions A.2 and A.5 impose restrictions on the structural parameters. Since

$$f(z,x)=f_\varepsilon(\widetilde{s}(z,x))f_x(x)\frac{\partial\widetilde{s}(z,x)}{\partial z},$$

where $\widetilde{s}(z,x)$ is the inverse supply function, Assumption A.2 will hold if we assume both f_x and f_ε satisfy the stated conditions on their respective domains and if $\widetilde{s}(z,x)$ is continuously differentiable of order $s'+1$. This requires U and P to be continuously differentiable of order $s'+2$. We will assume that these conditions hold.[5] Assumption A.5 will hold if the densities f_x and f_ε are bounded away from zero on $\widetilde{Z}\times\widetilde{X}$ and if the marginal utility function has nontrivial dependence on x. In any case, if there is concern that Assumption A.2 or Assumption A.5 may not be satisfied, since these are conditions on the density of observables, they can be tested.

The consistency and asymptotic normality of $\widehat{F}_\varepsilon$ and $\widehat{v}$ under Assumptions A.1–A.5 follow from Theorems 1 and 2 in Matzkin (2003). Let $\int K(x)^2=\int(\int K(s,x)\,ds)^2\,dx$, where $s\in\mathbf{R}$. Define C by

$$C=\left(P_{zz}\left(\frac{t_1}{x_1^*}\right)\right)^2\left(\frac{t_1}{(x_1^*)^2}\right)^2\left[\frac{dF_{Z|X=(x_1^*,x_2^*)}\left(\dfrac{t_1}{x_1^*}\right)}{dx_1}\right]^{-2}.$$

[4]See Aït-Sahalia (1994), Newey (1994), or Matzkin (2003).

[5]Conditions ensuring the differentiability of P are not available. See Ma, Trudinger, and Wang (2005), Loeper (2009), or Kim and McCann (2008) for recent results on smoothness of optimal maps in transportation problems.

THEOREM D.1: *Suppose that Assumptions A.1–A.6 are satisfied. Then* $\widehat{m}(t_1, t_2)$ *converges in probability to* $m(t_1, t_2)$ *and*

$$\sqrt{N\sigma_N^2}(\widehat{m}(t_1, t_2) - m(t_1, t_2)) \to N(0, V_m) \quad \textit{in distribution},$$

where

$$V_m = C\left\{\int K(x)^2\right\}\left(\frac{1}{f(\overline{x})} + \frac{1}{f(x^*)}\right)\bigl(F_{Z|X=\overline{x}}(\overline{z})\bigl(1 - F_{Z|X=\overline{x}}(\overline{z})\bigr)\bigr).$$

See Appendix F for the proof.

D.2. *Case 2*

We next consider the situation where the assumptions of Theorem 3.1 are satisfied. In this case, $x \in \mathbf{R}$ and we assume that for some unknown function m,

$$U_z(z, x, \varepsilon) = m(q(z, x), \varepsilon),$$

where $q : \mathbf{R}^2 \to \mathbf{R}$ is a *known* function. We assume that F_ε, the strictly increasing distribution of ε, is known. Then, as argued in the proof of Theorem 3.1, the derived supply function satisfies

$$s(x, e) = F^{-1}_{Z|X=x}(F_\varepsilon(e)).$$

This can be estimated by

$$\widehat{s}(x, e) = \widehat{F}^{-1}_{Z|X=x}(F_\varepsilon(e)),$$

where $\widehat{F}_{Z|X=x}$ is calculated as shown in the beginning of this section. Next, to estimate $m(t_1, t_2)$ at specified values t_1, t_2, let $\widehat{x}$ be such that

$$q(\widehat{s}(\widehat{x}, t_2), \widehat{x}) = t_1.$$

Then

$$\widehat{m}(t_1, t_2) = P_z(\widehat{s}(\widehat{x}, t_2)). \tag{D.2}$$

Theorem D.2 establishes the asymptotic properties of this estimator for the case where the function $q(z, x) = zx$ and P_z is known. This analysis serves as a prototype for more general cases. The assumptions of the theorem are very similar to those of Theorem D.1. Note that in this case, $n_x = 1$.

Let x^* be the value of x satisfying $q(s(x^*, t_2), x^*) = t_1$. In place of Assumptions A.5–A.6, we make the following assumptions:

ASSUMPTION A.5′: $x^* \neq 0$; $0 < f(x^*)$; there exist $\delta, \xi > 0$ such that $\forall x \in B(x^*, \xi), \forall \tilde{z} \in B(s(x, t_2), \xi)$, $f(x) \geq \delta$ and $f(\tilde{z}, x) \geq \delta$, and for all x in a neighborhood of x^*, $dF_{Z|X=x}(t_1/x)/dx \neq 0$.

ASSUMPTION A.6′: $t_1 \neq 0$ and it belongs to the interior of the support of $q(s(x, t_2), x)$.

The consistency and asymptotic normality of $\widehat{s}$ follow from Theorems 1 and 2 in Matzkin (2003).

THEOREM D.2: *Suppose that Assumptions A.1–A.4, and A.5′ and A.6′ are satisfied. Then $\widehat{m}(t_1, t_2)$ converges in probability to $m(t_1, t_2)$ and*

$$\sqrt{N\sigma_N}(\widehat{m}(t_1, t_2) - m(t_1, t_2)) \to N(0, V_{m'}) \quad \textit{in distribution},$$

where

$$V_{m'} = C\left\{\int K(x)^2\right\}\left(\frac{1}{f(x^*)}\right)\left(F_\varepsilon(t_2)(1 - F_\varepsilon(t_2))\right)$$

and

$$C = \left[P_{zz}\left(\frac{t_1}{x^*}\right)\right]^2 \left(\frac{t_1}{x^{*2}}\right)^2 \left[\frac{dF_{Z|X=x^*}(t_1/x^*)}{dx}\right]^{-2}.$$

See Appendix F for the proof.

The analysis for an estimator based on Theorem 3.2 is similar and for the sake of brevity is omitted. We next present some Monte Carlo evidence on the performance of these estimators.

APPENDIX E: PERFORMANCE OF THE ESTIMATORS

We present some limited Monte Carlo experiments that illustrate the performance of the estimation techniques presented in Appendix D. To obtain these results, we simulate data from a hedonic model using a range of parameter values. For each set of parameter values tested, we simulate 100 data sets, each with 500 observations. Then we estimate the marginal utility function using each of the data sets. We discuss the results of these simulations and present graphs which display the median estimates (across the 100 data sets) as well as the 5th and 95th percentile estimates. These results indicate that the techniques developed for estimating the nonadditive hedonic model work quite well.

E.1. *Model*

For workers, we assume that marginal utility is given by $U_z(z, x, \varepsilon) = m(q(z, x), \varepsilon)$, where $q(z, x) = zx$ and $m(q, \varepsilon) = \beta B q^{\beta-1}\varepsilon^{-\delta}$. For firms, we assume marginal productivity is given by $\Gamma(z, y, \eta) = Az^{\alpha}\eta$. We assume that $x \sim U[x_L, x_H]$, $\varepsilon \sim U[\varepsilon_L, \varepsilon_H]$, and $\eta \sim U[\eta_L, \eta_H]$. We assume that marginal utility depends on the known observable scale $q = zx$ and generate data from this model to estimate the function m. Parameter values used in our simulations are presented in Table I.

The baseline values in Table I are chosen to avoid numerical difficulties with parameter values near zero and to demonstrate the properties of the model. The alternative parameter values were chosen to examine the dependence of outcomes on model parameters. We examine the impact of variations in (x_U, β, δ). The parameter x_U affects the variance (and mean) of the observable variables and the size of the equilibrium support of (zx, ε). The parameters (β, δ) affect the degree of nonlinearity in the hedonic equilibrium, the shape of the hedonic pricing function, and, most importantly, the shape of the equilibrium support of (zx, ε). The features of the model and the equilibrium that have the most significant impact on the performance of the estimators are the relative variance of observables and unobserved variables and the equilibrium support of (zx, ε). As one would expect, increased variance of observables relative to unobservables reduces the sampling error of the estimator. Also, the estimator performs well on the interior of the support of (zx, ε), but less well near the boundary of the support where there are fewer observations.

The parameters $(x_L, \varepsilon_L, \varepsilon_U, \eta_L, \eta_U)$ have impacts that are qualitatively similar to the impacts of x_U. The parameter x_L affects the mean and variance of x and the equilibrium support of (zx, ε). The parameters $(\varepsilon_L, \varepsilon_U)$ affect the mean and variance of ε and the equilibrium support of (zx, ε). Increases

TABLE I

BASELINE PARAMETER VALUES AND ALTERNATIVE VALUES

Parameter Name	Baseline Values	Feasible Values	Alternative Values
x_L	1.0	$x_L > 0$	n.a.*
x_U	2.0	$x_U > x_L$	3.0
ε_L	1.0	$\varepsilon_L > 0$	n.a.
ε_U	2.0	$\varepsilon_U > \varepsilon_L$	n.a.
η_L	1.0	$\eta_L > 0$	n.a.
η_U	2.0	$\eta_U > \eta_L$	n.a.
α	0.25	$0 < \alpha < \beta$	n.a.
β	0.50	$\beta \neq 1$	0.75
δ	1.0	$\delta > 0$	2.0
A	1.0	$A > 0$	n.a.
B	1.0	$B > 0$	n.a.

*n.a. = not applicable because the specification in the column "Feasible Values" is maintained.

in the variance of ε reduce the precision of the estimates. The parameters (η_L, η_U) affect the equilibrium support of (zx, ε). For the sake of brevity, we do not report results for alternative values of α and for values of $\beta > 1$. These parameters affect the shape of the support of (zx, ε). In particular, when $\alpha < 1$ and $\beta > 1$, the support of (zx, ε) is confined to a very small region.[6]

E.2. *Estimation Results*

We use the procedure described in Section D.2 to estimate the supply function $z = s(x, \varepsilon)$. To estimate the model, we assume that $q(z, x) = zx$ and $\varepsilon \sim U[\varepsilon_L, \varepsilon_U]$. Under this assumption, we can compare the estimated values of $m(q, \varepsilon)$ with the true value. We estimate the marginal utility function $m(q, \varepsilon)$ for a selected set of values of q and ε in the relevant domain. The domain on which m is identified is both model-dependent and data-dependent. We illustrate this in our simulation results. The figures display the median values (across the 100 data sets) of our estimation results as well as the 5th and 95th percentiles.

Figure 3 presents results for the baseline model. The top two panels display the true function $m(q, \varepsilon)$ and the median of the estimates of that function. While m is well defined for all positive values of q and ε, the function is only identified on the funnel-shaped region underneath the graph in the figure. The limits of the region of identification are determined by the model. In particular, they are determined by the assumption that (x, ε, η) are each uniformly distributed. The shape of the region is determined in equilibrium and depends strongly on the supports of (x, ε, η) and on the curvature parameters (α, β, δ).

The figure shows that the median of the estimates of m are very accurate. The two functions in the top two panels are nearly identical. The bottom two panels show the estimated values of m for fixed values of ε and q, respectively. In these panels, the solid lines depict the true value of $m(q, \varepsilon)$, the dashed lines depict the median of the estimated values, the circles depict the 5th percentile estimates, and the plus signs depict the 95th percentile estimates. The solid lines and the dashed lines are indistinguishable. The 5th and 95th percentile values are also very close to the true values except near the boundaries of the supports. In the bottom left panel, the value of ε is fixed at 1.5. For this value of ε, the value of $m(q, \varepsilon)$ is accurately estimated for all values of $q \in [2, 24]$. The value of the function cannot be estimated for larger values of q. For other values of ε, the range of values of q that produce accurate estimates is different. In the bottom right panel, the variable q is fixed at 4.9564. For this value of q, $m(q, \varepsilon)$ is accurately estimated for values of ε ranging from about 1.3 to 1.9.

Figure 4 illustrates similar results when x_U is increased from 2.0 to 3.0. The precision of the estimates increases and the size of the region on which the function is identified increases. In Figure 3, the scale of the q axis ranges

[6]These results are available from the authors upon request.

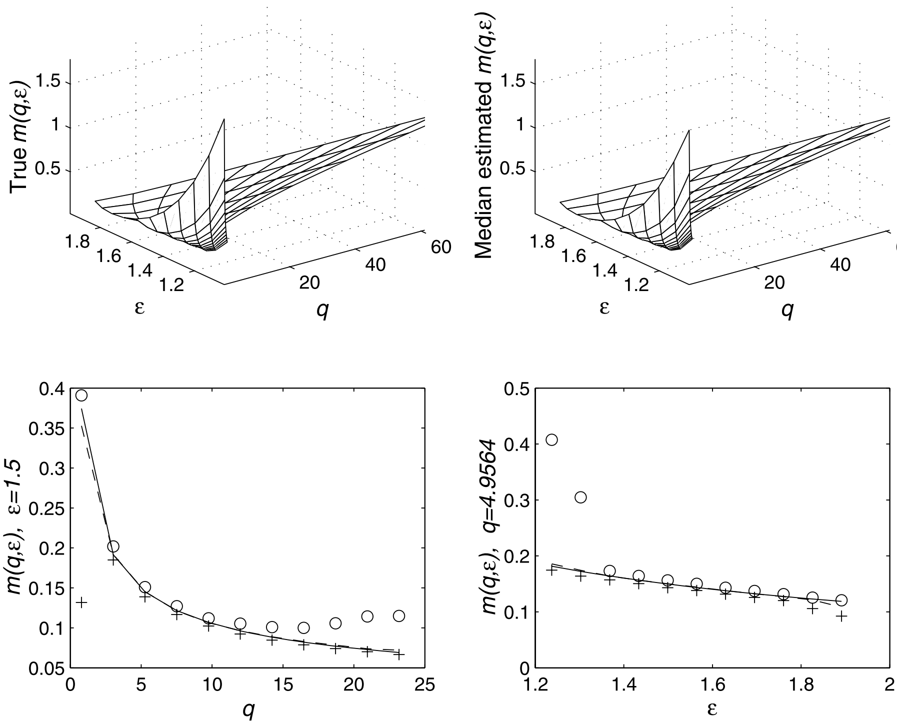

FIGURE 3.—Simulation results: baseline parameter values. The upper left panel plots the true values of $m(q,\varepsilon)$, where $q = zx$. The upper right panel plots the median of the estimates of $m(q,\varepsilon)$ (sample size 500; 100 Monte Carlo replications). The supports of the graphs indicate the equilibrium region on which the function m is identified. The lower left panel plots the true and estimated values of $m(q,\varepsilon)$ when $\varepsilon = 1.5$. The lower right panel plots the true and estimated values of $m(q,\varepsilon)$ when $q = 4.9564$. The solid lines plot the true function values, the dashed lines plot the medians of the estimated values, the circles plot the 95th percentile estimates, and the plus symbols plot the 5th percentile estimates. True baseline parameter values are given in Table I.

from 0 to 60. In contrast, in Figure 4, the q axis scale ranges from 0 to 150. In both Figures 3 and 4, the function m is accurately estimated for all values of $\varepsilon \in [1.2, 1.8]$ when q is small. However, when q is large, the interval in the ε dimension within which m can be accurately estimated is smaller.

Figure 5 illustrates the impact of increasing β to 0.75. This change has a dramatic impact on the support of (zx, ε) and hence on the region on which m is identified. The scale of the q axis in Figure 5 ranges from 0 to 2.5. Within this range, m can be estimated accurately, but the equilibrium provides no information about the model for values of q outside this region. As β approaches 1, the performance of the estimator deteriorates. In the limiting case where $\beta = 1$, x does not affect marginal utility.

Finally, Figure 6 illustrates the impact of increasing δ to 2.0. This change drastically increases the equilibrium support of z and hence of (zx, ε). Notice

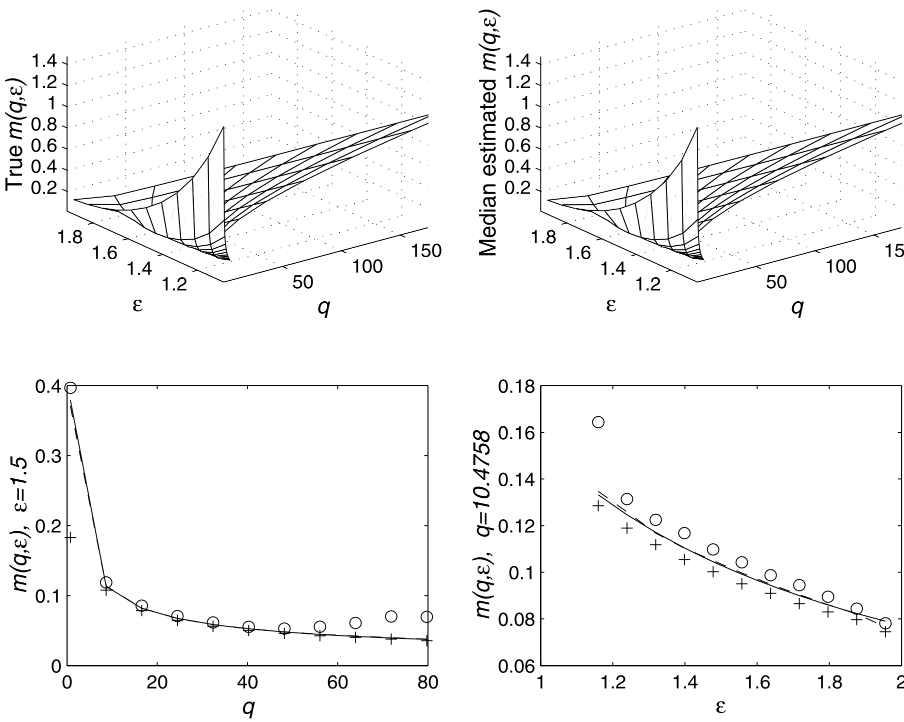

FIGURE 4.—Simulation results: $x_U = 3.0$. The upper left panel plots the true values of $m(q, \varepsilon)$, where $q = zx$. The upper right panel plots the median of the estimates of $m(q, \varepsilon)$ (sample size 500; 100 Monte Carlo replications). The supports of the graphs indicate the equilibrium region on which the function m is identified. The lower left panel plots the true and estimated values of $m(q, \varepsilon)$ when $\varepsilon = 1.5$. The lower right panel plots the true and estimated values of $m(q, \varepsilon)$ when $q = 10.4758$. The solid lines plot the true function values, the dashed lines plot the medians of the estimated values, the circles plot the 95th percentile estimates, and the plus symbols plot the 5th percentile estimates. All true parameter values except x_U are identical to the baseline parameter values. This case used the value $x_U = 3.0$.

that the scale of the q axis in Figure 6 ranges from 0 to 500. The upper right panel of Figure 6 shows that the median of the estimates of m is very similar to the true value of m (depicted in the upper left panel). The lower left panel shows that when $\varepsilon = 1.5$, the value of m is accurately estimated for values of q ranging from 10 to 80. Similarly, the lower right shows that the value of m is accurately estimated when $q = 25.07$ for all values of ε ranging from 1.4 to 1.7.

These figures illustrate that the estimator performs well in the interior of the support of (zx, ε). Remember that we estimate the supply function $z = s(x, \varepsilon)$ and then use the estimated function, the marginal price function $P_z(z)$, and knowledge of the index structure $U_z(z, x, \varepsilon) = m(q, \varepsilon)$, where $q = zx$, to estimate m. Crucial determinants of the performance of the estimator of m

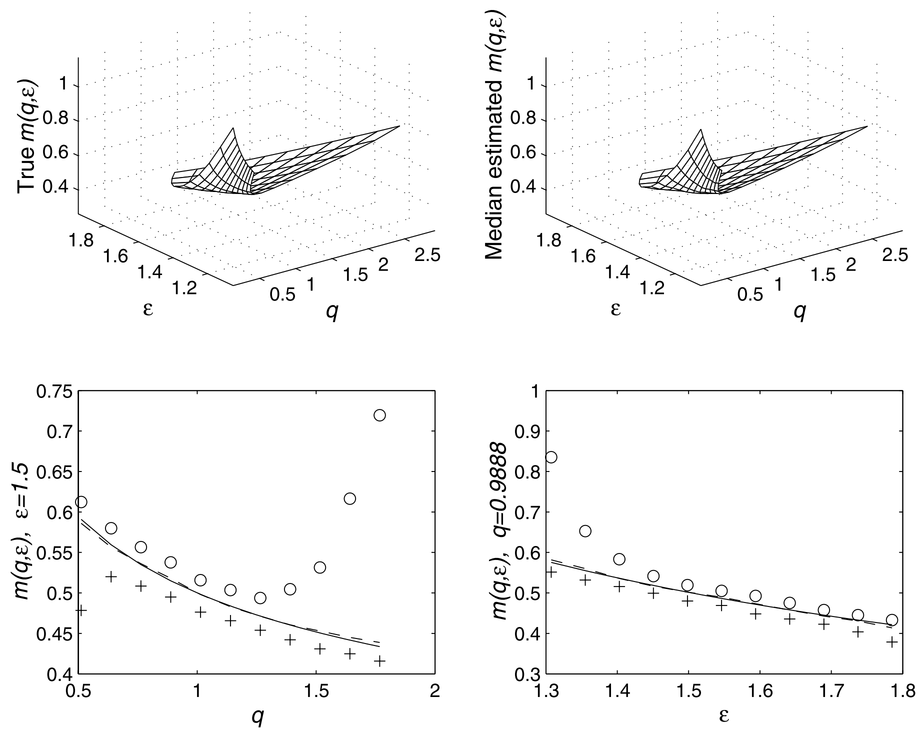

FIGURE 5.—Simulation results: $\beta = 0.75$. The upper left panel plots the true values of $m(q, \varepsilon)$, where $q = zx$. The upper right panel plots the median of the estimates of $m(q, \varepsilon)$ (sample size 500; 100 Monte Carlo replications). The supports of the graphs indicate the equilibrium region on which the function m is identified. The lower left panel plots the true and estimated values of $m(q, \varepsilon)$ when $\varepsilon = 1.5$. The lower right panel plots the true and estimated values of $m(q, \varepsilon)$ when $q = 0.9888$. The solid lines plot the true function values, the dashed lines plot the medians of the estimated values, the circles plot the 95th percentile estimates, and the plus symbols plot the 5th percentile estimates. All true parameter values except β are identical to the baseline parameter values. This case used the value $\beta = 0.75$.

are the relative variance of observables and unobservables and the equilibrium support of (zx, ε). In applications, since researchers must first estimate $z = s(x, \varepsilon)$, the first stage estimate can be used to construct a residual for each observation and to estimate the joint density of (zx, ε). This joint density provides information as to the region in (zx, ε) where many observations are available and where it is possible to estimate m accurately.

APPENDIX F: PROOFS

PROOF OF THEOREM D.1: We apply the delta method developed in Newey (1994). Let $F(z, x)$ denote the c.d.f. of the vector of observable variables

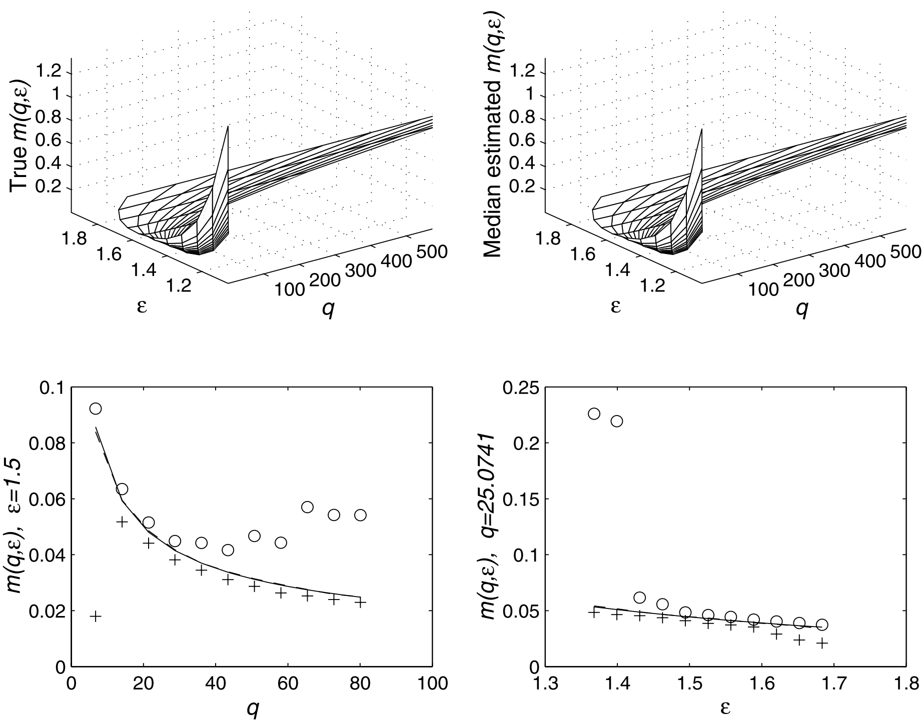

FIGURE 6.—Simulation results: $\delta = 2.0$. The upper left panel plots the true values of $m(q, \varepsilon)$, where $q = zx$. The upper right panel plots the median of the estimates of $m(q, \varepsilon)$ (sample size 500; 100 Monte Carlo replications). The supports of the graphs indicate the equilibrium region on which the function m is identified. The lower left panel plots the true and estimated values of $m(q, \varepsilon)$ when $\varepsilon = 1.5$. The lower right panel plots the true and estimated values of $m(q, \varepsilon)$ when $q = 25.0741$. The solid lines plot the true function values, the dashed lines plot the medians of the estimated values, the circles plot the 95th percentile estimates, and the plus symbols plot the 5th percentile estimates. All true parameter values except δ are identical to the baseline parameter values. This case used the value $\delta = 2.0$.

(Z, X), let $f(z, x)$ denote its p.d.f., let $f(x)$ denote the marginal p.d.f. of X, and let $F_{Z|X=x}$ denote the conditional c.d.f. of Z given $X = x$. Recall that $\widetilde{Z} \times \widetilde{X}$ is the compact support of (Z, X). Let $L = 3$ be the dimension of $\widetilde{Z} \times \widetilde{X}$. For any function $G : \mathbf{R}^L \to \mathbf{R}$, define $g(z, x) = \partial^L G(z, x)/\partial z\, \partial x$, $g(x) = \int_{-\infty}^{\infty} g(s, x)\, ds$, $G_{Z|X=x}(z') = (\int_{-\infty}^{z'} g(s, x)\, ds)/g(x)$, and $\widetilde{G}_Z(z, x) = \int_{-\infty}^{z} g(s, x)\, ds = \int_{-\infty}^{\infty} 1[s \le z] g(s, x)\, ds$, where $1[\cdot] = 1$ if $[\cdot]$ is true and equals zero otherwise. Let $\mathcal{C}$ denote a compact set in $\mathbf{R}^L$ that strictly includes $\widetilde{Z} \times \widetilde{X}$. Let B denote the set of all functions $G : \mathbf{R}^L \to \mathbf{R}$ such that $g(z, x)$ has bounded, continuous derivatives up to order 3 and vanishes outside $\mathcal{C}$. Let $\widetilde{B}$ denote the set of all functions $\widetilde{G}_Z$ that are derived from some G in B. Since there is a 1–1 relationship between functions in B and functions in $\widetilde{B}$, we can define a functional on B or on $\widetilde{B}$ without altering its definition. Let $\|G\|$ denote the maximum of the sup norms of $g(z,x)$ and its derivatives up to third order. If $H \in B$, there exists $\rho_1 > 0$ such that if $\|H\| \le \rho_1$, then for some $0 < a, b < \infty$, all x in a neighborhood of (x_1^*, x_2^*), and all $\widetilde{z} \in B(v(x_1,t_2),\xi)$, $|h(x)| \le a\|H\|$, $|\int^{\widetilde{z}} h(s,x)\,ds| \le a\|H\|$, $f(x)+h(x) \ge b|f(x)|$, and $f(\widetilde{z},x) + h(\widetilde{z},x) \ge b|f(\widetilde{z},x)|$. Let $\overline{x} = (\overline{x}_1, \overline{x}_2)$.

We first derive the asymptotic behavior of $\widehat{x}_1^*$, defined as the value of x_1 that, given $\widehat{v}$, satisfies

$$q_1(\widehat{v}(\widehat{x}_1^*, t_2), \widehat{x}_1^*) = t_1.$$

Recall that $x_2^* = t_2 + e^*$ and

$$\widehat{v}(x_1, t_2) = \widehat{F}^{-1}_{Z|X=(x_1,x_2^*)}\big(\widehat{F}_{Z|X=\overline{x}}(\overline{z})\big).$$

Hence, $\widehat{x}_1^*$ satisfies

$$\widehat{F}^{-1}_{Z|X=(\widehat{x}_1^*,x_2^*)}\big(\widehat{F}_{Z|X=\overline{x}}(\overline{z})\big) \cdot \widehat{x}_1^* = t_1$$

or

$$\widehat{F}_{Z|X=\overline{x}}(\overline{z}) = \widehat{F}_{Z|X=(\widehat{x}_1^*,x_2^*)}\left(\frac{t_1}{\widehat{x}_1^*}\right).$$

Analogously, the population value x_1^* satisfies

$$F_{Z|X=\overline{x}}(\overline{z}) = F_{Z|X=(x_1^*,x_2^*)}\left(\frac{t_1}{x_1^*}\right).$$

Define the functional $\rho(G, x_1)$ on $B \times \widetilde{X}_1$ by

$$\rho(G, x_1) = G_{Z|X=\overline{x}}(\overline{z}) - G_{Z|X=(x_1,x_2^*)}\left(\frac{t_1}{x_1}\right).$$

For any x_1 in a small enough neighborhood of x_1^*, any G in a small enough neighborhood of F, any H such that $\|H\|$ is small enough, and any Δx_1 such that $|\Delta x_1|$ is small enough,

$$\rho(G+H, x_1) - \rho(G, x_1)$$
$$= \frac{\int_{-\infty}^{\overline{z}} (g(s,\overline{x}) + h(s,\overline{x}))\,ds}{(g(\overline{x}) + h(\overline{x}))}$$

$$
\begin{aligned}
&- \frac{\int_{-\infty}^{\overline{z}} g(s,\overline{x})\,ds}{g(\overline{x})} \\
&- \Bigg[\frac{\int_{-\infty}^{(t_1/x_1)} (g(s,x_1,x_2^*) + h(s,x_1,x_2^*))\,ds}{(g(x_1,x_2^*) + h(x_1,x_2^*))} \\
&- \frac{\int_{-\infty}^{(t_1/x_1)} g(s,x_1,x_2^*)\,ds}{g(x_1,x_2^*)}\Bigg] \\
&= \frac{g(\overline{x})\int_{-\infty}^{\overline{z}} h(s,\overline{x})\,ds}{g(\overline{x})(g(\overline{x}) + h(\overline{x}))} - \frac{h(\overline{x})\int_{-\infty}^{\overline{z}} g(s,\overline{x})\,ds}{g(\overline{x})(g(\overline{x}) + h(\overline{x}))} \\
&- \Bigg[\frac{g(x_1,x_2^*)\int_{-\infty}^{(t_1/x_1)} h(s,x_1,x_2^*)\,ds}{g(x_1,x_2^*)(g(x_1,x_2^*) + h(x_1,x_2^*))} \\
&- \frac{h(x_1,x_2^*)\int_{-\infty}^{(t_1/x_1)} g(s,x_1,x_2^*)\,ds}{g(x_1,x_2^*)(g(x_1,x_2^*) + h(x_1,x_2^*))}\Bigg] \\
&= \frac{\int_{-\infty}^{\overline{z}} h(s,\overline{x})\,ds - h(\overline{x})G_{Z|X=\overline{x}}(\overline{z})}{g(\overline{x})} \\
&- \frac{\int_{-\infty}^{(t_1/x_1)} h(s,x_1,x_2^*)\,ds - h(x_1,x_2^*)G_{Z|X=(x_1,x_2^*)}(t_1/x_1)}{g(x_1,x_2^*)} \\
&- \frac{\int_{-\infty}^{\overline{z}} h(s,\overline{x})\,ds - h(\overline{x})G_{Z|X=\overline{x}}(\overline{z})}{g(\overline{x})}\frac{h(\overline{x})}{g(\overline{x}) + h(\overline{x})} \\
&+ \frac{\Big[\int_{-\infty}^{(t_1/x_1)} h(s,x_1,x_2^*)\,ds - h(x_1,x_2^*)G_{Z|X=(x_1,x_2^*)}(t_1/x_1)\Big]}{g(x_1,x_2^*)} \\
&\times \frac{h(x_1,x_2^*)}{g(x_1,x_2^*) + h(x_1,x_2^*)}.
\end{aligned}
$$

Define

$$D_F\rho(G,x_1;H) = \frac{\int_{-\infty}^{\overline{z}} h(s,\overline{x})\,ds - h(\overline{x})G_{Z|X=\overline{x}}(\overline{z})}{g(\overline{x})} - \frac{\int_{-\infty}^{(t_1/x_1)} h(s,x_1,x_2^*)\,ds - h(x_1,x_2^*)G_{Z|X=(x_1,x_2^*)}(t_1/x_1)}{g(x_1,x_2^*)}$$

and

$$R_F\rho(G,x_1;H) = -\left[\frac{\int_{-\infty}^{\overline{z}} h(s,\overline{x})\,ds - h(\overline{x})G_{Z|X=\overline{x}}(\overline{z})}{g(\overline{x})}\right]\frac{h(\overline{x})}{g(\overline{x})+h(\overline{x})} + \left[\frac{\int_{-\infty}^{(t_1/x_1)} h(s,x_1,x_2^*)\,ds - h(x_1,x_2^*)G_{Z|X=(x_1,x_2^*)}(t_1/x_1)}{g(x_1,x_2^*)}\right] \times \frac{h(x_1,x_2^*)}{g(x_1,x_2^*)+h(x_1,x_2^*)}.$$

Then, for some $a_1 < \infty$, and all (G,x) in the small neighborhood of (F,x^*), as described, we have that

$$|D_F\rho(G,x_1;H)| \le a_1\|H\|, \quad |R_F\rho(G,x_1;H)| \le a_1\|H\|^2$$

and

$$\rho(G+H,x_1) - \rho(G,x_1) = D_F\rho(G,x_1;H) + R_F\rho(G,x_1;H).$$

It follows that $D_F\rho(G,x_1;H)$ is the Fréchet derivative of ρ with respect to F when $F=G$. Also, for such (G,x) as described above,

$$\rho(G,x_1+\Delta x_1) - \rho(G,x_1) = -G_{Z|X=(x_1+\Delta x_1,x_2^*)}\left(\frac{t_1}{x_1+\Delta x_1}\right) + G_{Z|X=(x_1,x_2^*)}\left(\frac{t_1}{x_1}\right) = -\frac{\int_{-\infty}^{t_1/(x_1+\Delta x_1)} g(s,x_1+\Delta x_1,x_2^*)\,ds}{g(x_1+\Delta x_1,x_2^*)} + \frac{\int_{-\infty}^{t_1/x_1} g(s,x_1,x_2^*)\,ds}{g(x_1,x_2^*)}$$

$$= -\frac{\int_{-\infty}^{t_1/(x_1+\Delta x_1)} \left(g(s, x_1, x_2^*) + \frac{\partial g(s, x_1, x_2^*)}{\partial x_1} \Delta x_1 + R_{g,1} \right) ds}{g(x_1 + \Delta x_1, x_2^*)} + \frac{\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds}{g(x_1, x_2^*)},$$

where the last equality follows by Taylor's theorem with $|R_{g,1}| \le a_2 |\Delta x_1|^2$ for some $a_2 < \infty$ which depends on (g, x_1, x_2^*). Again using Taylor's theorem, it follows that for some $a_3 < \infty$, which depends on (g, x_1, x_2^*), and for $R_{g,2}$ and $R_{g,3}$ with $|R_{g,2}| \le a_3 |\Delta x_1|^2$ and $|R_{g,3}| \le a_3 |\Delta x_1|^2$,

$$\begin{aligned} &\rho(G, x_1 + \Delta x_1) - \rho(G, x_1) \\ &= -\left[\int_{-\infty}^{t_1/x_1} \left(g(s, x_1, x_2^*) + \frac{\partial g(s, x_1, x_2^*)}{\partial x_1} \Delta x_1 + R_{g,1} \right) ds \right] g(x_1, x_2^*) \\ &\quad \Big/ (g(x_1 + \Delta x_1, x_2^*) g(x_1, x_2^*)) \\ &\quad - \left[g\left(\frac{t_1}{x_1}, x_1, x_2^* \right) \left(\frac{-t_1}{x_1^2} \right) \Delta x_1 + \frac{\partial g\left(\frac{t_1}{x_1}, x_1, x_2^* \right)}{\partial x_1} \left(\frac{-t_1}{x_1^2} \right) (\Delta x_1)^2 + R_{g,3} \right] g(x_1, x_2^*) \\ &\quad \Big/ (g(x_1 + \Delta x_1, x_2^*) g(x_1, x_2^*)) \\ &\quad + \frac{\left[\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds \right] \left[g(x_1, x_2^*) + \frac{\partial g(x_1, x_2^*)}{\partial x_1} \Delta x_1 + R_{g,2} \right]}{g(x_1 + \Delta x_1, x_2^*) g(x_1, x_2^*)}. \end{aligned}$$

Hence,

$$\begin{aligned} &\rho(G, x_1 + \Delta x_1) - \rho(G, x_1) \\ &= -\left[g(x_1, x_2^*) \int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds + g(x_1, x_2^*) \int_{-\infty}^{t_1/x_1} \frac{\partial g(s, x_1, x_2^*)}{\partial x_1}\, ds\, \Delta x_1 \right] \end{aligned}$$

$$
\begin{aligned}
&\Big/ (g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)) \\
&- \frac{\left[\int_{-\infty}^{t_1/x_1} R_{g,1}\, ds\right] g(x_1, x_2^*)}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)} \\
&- \frac{\left[g(x_1, x_2^*) g\left(\frac{t_1}{x_1}, x_1, x_2^*\right)\left(\frac{-t_1}{x_1^2}\right)\Delta x_1\right]}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)} \\
&- \frac{\left[\frac{\partial g\left(\frac{t_1}{x_1}, x_1, x_2^*\right)}{\partial x_1}\left(\frac{-t_1}{x_1^2}\right)(\Delta x_1)^2 + R_{g,3}\right] g(x_1, x_2^*)}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)} \\
&+ \left[g(x_1, x_2^*)\left[\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds\right]\right. \\
&\left. + \frac{\partial g(x_1, x_2^*)}{\partial x_1}\left[\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds\right]\Delta x_1\right] \\
&\Big/ (g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)) \\
&+ \frac{\left[\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds\right][R_{g,2}]}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)}.
\end{aligned}
$$

Let

$$
\begin{aligned}
&D_{x_1}\rho(G, x_1; \Delta x_1) \\
&\quad = -\frac{\int_{-\infty}^{t_1/x_1} \frac{\partial g(s, x_1, x_2^*)}{\partial x_1}\, ds\, \Delta x_1}{g(x_1, x_2^*)} - \frac{g\left(\frac{t_1}{x_1}, x_1, x_2^*\right)\left(\frac{-t_1}{x_1^2}\right)\Delta x_1}{g(x_1, x_2^*)} \\
&\qquad + \frac{\frac{\partial g(x_1, x_2^*)}{\partial x_1}\left[\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds\right]\Delta x_1}{g(x_1, x_2^*)^2}
\end{aligned}
$$

and

$$
\begin{aligned}
&R_{x_1}\rho(G, x_1; \Delta x_1) \\
&\quad = -D_{x_1}\rho(G, x_1; \Delta x_1)\left[\frac{g(x_1+\Delta x_1, x_2^*) - g(x_1, x_2^*)}{g(x_1+\Delta x_1, x_2^*)}\right]
\end{aligned}
$$

$$- \frac{\left[\int_{-\infty}^{t_1/x_1} R_{g,1}\, ds\right] g(x_1, x_2^*)}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)}$$

$$- \frac{\left[\frac{\partial g\left(\frac{t_1}{x_1}, x_1, x_2^*\right)}{\partial x_1}\left(\frac{-t_1}{x_1^2}\right)(\Delta x_1)^2 + R_{g,3}\right] g(x_1, x_2^*)}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)}$$

$$+ \frac{\left[\int_{-\infty}^{t_1/x_1} g(s, x_1, x_2^*)\, ds\right][R_{g,2}]}{g(x_1+\Delta x_1, x_2^*) g(x_1, x_2^*)}.$$

Then, for some $a_4 < \infty$, and all (G, x_1) in a neighborhood of (F, x_1^*)

$$|D_{x_1}\rho(G, x_1; \Delta x_1)| \leq a_4|\Delta x_1|, \quad |R_{x_1}\rho(G, x_1; \Delta x_1)| \leq a_4|\Delta x_1|^2$$

and

$$\rho(G, x_1 + \Delta x_1) - \rho(G, x_1) = D_{x_1}\rho(G, x_1; \Delta x_1) + R_{x_1}\rho(G, x_1; \Delta x_1).$$

It follows that $D_{x_1}\rho(G, x_1; \Delta x_1)$ is the Fréchet derivative of ρ with respect to x_1 at (G, x_1).

In a similar way, one can show that each of the Fréchet derivatives $D_F\rho(G, x_1; H)$ and $D_{x_1}\rho(G, x_1; \Delta x_1)$ is itself Fréchet differentiable with respect to G and x_1, and, moreover, our assumptions imply that these Fréchet derivatives are continuous and uniformly bounded on a neighborhood of (F, x_1^*). Since $\rho(F, x_1^*) = 0$ and, by our assumptions, the continuous $D_{x_1}\rho(G, x; \Delta x_1)$ is invertible in such neighborhood, it follows by the Implicit Function Theorem of Hildebrandt and Graves (1927) (see Zeidler (1991, p. 150)), that there exists a unique functional $\kappa(G)$ on a neighborhood of F such that for all G in that neighborhood,

$$\rho(G, \kappa(G)) = 0.$$

Moreover, since ρ is a C^2 map, κ is also a C^2 map on such a neighborhood. For any G in the neighborhood of F, the Fréchet derivative $D_F\kappa(G; H)$ of κ at G in the direction H satisfies

$$D_{x_1}\rho(G, \kappa(G); D_F\kappa(G; H)) = -D_F\rho(G, \kappa(G); H).$$

Hence, since

$$D_{x_1}\rho(G, x_1; \Delta x_1) = -\left[\frac{dG_{Z|X=(x_1,x_2^*)}\left(\frac{t_1}{x_1}\right)}{dx_1}\right]\Delta x_1,$$

the Fréchet derivative of κ is given by

$$D_F\kappa(G;H) = \left[\frac{dG_{Z|X=(\kappa(G),x_2^*)}\left(\frac{t_1}{\kappa(G)}\right)}{dx_1}\right]^{-1}\bigl[D_F\rho(G,\kappa(G);H)\bigr].$$

In particular,

$$D_F\kappa(F;H) = \left[\frac{dF_{Z|X=(x_1^*,x_2^*)}\left(\frac{t_1}{x_1^*}\right)}{dx_1}\right]^{-1}[D_F\rho(F,x_1^*;H)].$$

The uniform boundedness of the second and first order Fréchet derivatives of ρ in a neighborhood of (F, x_1^*) implies by Taylor's theorem in Banach spaces that the remainder $R_F\kappa(F;H) = \kappa(F+H) - \kappa(F) - D_F\kappa(F;H)$ satisfies

$$|R_F\kappa(F;H)| \leq c\|H\|^2$$

for some $c < \infty$. Letting $H = \widehat{F} - F$, it follows by our assumptions and the Delta method in Newey (1994) that

$$\sqrt{N\sigma_N^2}(\widehat{x}_1^* - x_1^*) = \sqrt{N\sigma_N^2}(\kappa(\widehat{F}) - \kappa(F)) \to N(0, V_{\widehat{x}}),$$

where

$$V_{\widehat{x}} = \left[\frac{dF_{Z|X=(x_1^*,x_2^*)}\left(\frac{t_1}{x_1^*}\right)}{dx_1}\right]^{-2}\left\{\int K(x)^2\right\} \times \left(\frac{1}{f(\overline{x})} + \frac{1}{f(x^*)}\right)\bigl(F_{Z|X=\overline{x}}(\overline{z})\bigl(1 - F_{Z|X=\overline{x}}(\overline{z})\bigr)\bigr).$$

Since

$$\widehat{m}(t_1,t_2) = P_z(\widehat{v}(\widehat{x}_1^*,t_2)) = P_z\left(\frac{t_1}{\widehat{x}_1^*}\right),$$

it follows by the standard Delta method that

$$\sqrt{N\sigma_N^2}(\widehat{m}(t_1,t_2) - m(t_1,t_2)) \to N(0, V_m),$$

where

$$V_m = C\left\{\int K(x)^2\right\}\left(\frac{1}{f(\overline{x})} + \frac{1}{f(x^*)}\right)\bigl(F_{Z|X=\overline{x}}(\overline{z})\bigl(1 - F_{Z|X=\overline{x}}(\overline{z})\bigr)\bigr)$$

and

$$C = \left(P_{zz}\left(\frac{t_1}{x_1^*}\right)\right)^2 \left(\frac{t_1}{(x_1^*)^2}\right)^2 \left[\frac{dF_{Z|X=(x_1^*,x_2^*)}\left(\frac{t_1}{x_1^*}\right)}{dx_1}\right]^{-2}.$$

Q.E.D.

PROOF OF THEOREM D.2: The method of proof is very similar to that of Theorem D.1. The only difference is that $\widehat{F}_{Z|X=\overline{x}}(\overline{z})$ and $F_{Z|X=\overline{x}}(\overline{z})$ in the proof of Theorem D.1 are now replaced by $F_\varepsilon(t_2)$. Following the same steps as in the proof of Theorem D.1, it is then easy to show that

$$\sqrt{N\sigma_N}(\widehat{m}(t_1,t_2) - m(t_1,t_2)) \to N(0, V_{m'}),$$

where

$$V_{m'} = C\left\{\int K(x)^2\right\}\left(\frac{1}{f(x^*)}\right)\big(F_\varepsilon(t_2)(1-F_\varepsilon(t_2))\big)$$

and C is as in the proof of Theorem D.1. *Q.E.D.*

REFERENCES

AÏT-SAHALIA, Y. (1994): "The Delta and Bootstrap Methods for Nonlinear Functionals of Nonparametric Kernel Estimators Based on Dependent Multivariate Data," Unpublished Manuscript, Princeton University. [11]

GALE, D., and H. NIKAIDO (1965): "The Jacobian Matrix and Global Univalence of Mappings," *Mathematische Annalen*, 159, 81–93.

HILDEBRANDT, T. H., and L. M. GRAVES (1927): "Implicit Functions and Their Differentials in General Analysis," *Transactions of the American Mathematical Society*, 29, 127–153. [25]

KIM, Y. H., and R. MCCANN (2008): "Towards the Smoothness of Optimal Maps on Riemannian Submersions and Riemannian Products (of Round Spheres in Particular)," Unpublished Manuscript, Department of Mathematics, University of Toronto. Available at http://www.math.toronto.edu/mccann/papers/RiemSub.pdf. [11]

LOEPER, G. (2009): "On the Regularity of Solutions of Optimal Transportation Problems," *Acta Mathematica*, 202, 241–283. [11]

MA, X.-N., N. S. TRUDINGER, and X.-J. WANG (2005): "Regularity of Potential Functions of the Optimal Transportation Problem," *Archive for Rational Mechanics and Analysis*, 177, 151–183. [11]

MATZKIN, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 71, 1339–1375. [10,11,13]

NEWEY, W. K. (1994): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253. [11,18,26]

TINBERGEN, J. (1956): "On the Theory of Income Distribution," *Weltwirtschaftliches Archiv*, 77, 155–173. [6]

ZEIDLER, E. (1991): *Nonlinear Functional Analysis and Its Applications*. New York: Springer-Verlag. [25]

Dept. of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, U.S.A., and University College Dublin, American Bar Foundation and Cowles Foundation, Yale University; jjh@uchicago.edu, *Dept. of Economics, University of California, Los Angeles, 8283 Bunche Hall, Los Angeles, CA 90095, U.S.A.; matzkin@econ.ucla.edu,*

and

Dept. of Economics, University College London, Gower Street, London, WC1E 6BT, U.K. and CEMMAP and Institute for Fiscal Studies; l.nesheim@ucl.ac.uk.

Manuscript received March, 2006; final revision received April, 2009.