SEMIPARAMETRIC INSTRUMENTAL VARIABLES ESTIMATION AND ITS APPLICATION TO DYNAMIC OLIGOPOLY

SANGIN PARK*

Department of Economics
SUNY at Stony Brook
Stony Brook, NY 11794-4384
U.S.A.
Email: sanpark@notes.cc.sunysb.edu

This Draft: August 1999

This paper considers a semiparametric regression model in which the error term is correlated with the nonparametric part. A technical difficulty of this semiparametric regression model is that we can not eliminate the nonparametric part in the two-step estimation procedure of a typical semiparametric regression model. Yet, we can still obtain a semiparametric estimator, called a semiparametric instrumental variables (SIV) estimator, with consistency and asymptotic normality if there exist two sets of instrumental variables (IVs) satisfying both an identification condition and an orthogonality condition. The paper provides two generic examples in which we can construct these two sets of IVs, and then discusses an empirical example of the application of the SIV estimation procedure to estimate network effects in the U.S. home VCR market. This empirical example indicates that the SIV estimation procedure can be applied to some structural models of dynamic oligopoly in order to avoid a prohibitive computational burden of calculating each firm's value functions in equilibrium for each candidate value of the parameter vector.

KEYWORDS: semiparametric instrumental variables estimation, filtering, identification, orthogonality, dynamic oligopoly, network externality.

JEL Classification: C14; C51; D43.

* I am grateful for helpful comments from Ariel Pakes, Don Andrews, Steven Berry, Oliver Linton, John Rust, Lanier Benkard, Jinyong Hahn, Soiliou Namoro, and Ito Harumi. The previous version of this paper has been presented at the 1997 Winter Meeting of North America Econometric Society, Ewha University, SUNY at Stony Brook, Universite de Toulouse I, and Yale University. All errors and omissions are my own.

1. INTRODUCTION

A COMMON ASSUMPTION IN SEMIPARAMETRIC REGRESSION MODELS is that the error term, say $u_j \in R$, is mean independent of the nonparametric part, say $\varphi_0(\cdot)$, where φ_0 is an unknown real function.[1] Then a typical semiparametric regression model is specified as follows:

(1.1) $\quad y_j = f(x_j;\ \theta_0) + \varphi_0(v_j) + u_j,$

with

(1.2) $\quad E[u_j \mid v_j] = 0,$

where $(y_j, x_j', v_j')'$ is an $R \times R^d \times R^l$-valued vector of random variables, θ_0 is a R^k-valued vector of parameters which we want to estimate, and f is a known function from $R^d \times R^k$ to R. Since $\varphi_0(\cdot)$ is unknown, $\varphi_0(v_j)$ cannot be observed.

Usually, the conditional moment restriction (CMR) in (1.2) is applied to eliminate the nonparametric part $\varphi_0(v_j)$,[2] leading to:[3]

(1.3) $\quad y_j - E[y_j \mid v_j] = f(x_j;\ \theta_0) - E[f(x_j;\ \theta_0) \mid v_j] + u_j.$

In a two-step estimation procedure, the conditional expectations, $E[y_j \mid v_j]$ and $E[f(x_j;\ \theta) \mid v_j]$ for any θ, are estimated nonparametrically in the first step. In the second step, if x_j is also mean independent of u_j, we can obtain a semiparametric least squares estimator for θ_0 (Robinson

---

[1] All the functions in the paper are assumed to be measurable functions.

[2] In the case of partially linear regression models, the nonparametric part may be approximated by some series estimates. Refer to Andrews (1991) for the regularity conditions for this case. If the density of the data $(y_j, x_j', v_j')'$ satisfies an index restriction, we may obtain a semiparametric maximum likelihood estimator for θ_0 (see Ai (1997)).

(1988)), or a semiparametric nonlinear least squares estimator (Andrews (1994)). If x_j is correlated with u_j, and if there is an instrumental vector, we may obtain a semiparametric method of moments estimator as a special case of Pakes and Olley (1995).

In this paper, we consider a case in which an error term, say $\xi_j \in R$, is correlated with a nonparametric part, say $\phi(v_j)$, where ϕ is an unknown function from R^l to R. The regression model this paper considers takes the form:

(1.4) $y_j = f(x_j; \theta_0) + \phi(v_j) + \xi_j$,

with

(1.5) $E[\xi_j \mid v_j] \neq 0$.[4]

An example of the regression model in (1.4) and (1.5) can be derived from structural models of dynamic oligopoly. Dynamic oligopoly is a situation in which firms' price-settings (or quantity-settings) are strategically interdependent and have durable effects on the stream of their profits. Dynamic oligopoly fits many industries characterized by the significance of network externalities, learning-by-doing, or informational product differentiation. For example, semiconductor, especially DRAM, manufacturing is considered as a dynamic oligopoly (see Scherer (1996)). Learning-by-doing is significant in the semiconductor industry since a firm's current output level affects its future production costs and thus its future profits. Particularly in the DRAM market, large investments in R&D and learning-by-doing lead to a fairly concentrated industry, and strategic interactions are significant and change dramatically over the product

[3] The conditional moment restriction in (1.2) implies: $E[y_j|v_j] = E[f(x_j;\theta_0)|v_j] + \varphi_0(v_j)$. Subtracting this from (1.1) yields (1.3).

[4] The estimation procedure of this paper can be extended to a more general case:

$\xi_j = f(y_j, x_j; \theta_0) + \phi(v_j)$, with $E[\xi_j \mid v_j] \neq 0$.

cycle.[5] Dynamic oligopoly is also a relevant market structure of many industries characterized by the significance of network externalities (see Park (1999)). In the presence of network externalities, a firm's current price may affect its future network size and thus future profits and survival. Network externalities are significant in such industries as the computer industry, the broadcasting industry, and some consumer electronics industries. In most of these industries, technologies are sponsored and markets are oligopolistic.[6] Another example of dynamic oligopoly includes industries characterized by informational product differentiation. In informational product differentiation, a firm's current output level can affect its future profits since the quality of a product is learned by consumption experience. In the pharmaceutical industry, a branded drug usually enjoys informational product differentiation over generics (see Schmalensee (1982); Currie and Park (1999)).

To my knowledge, however, there is no tractable estimation procedure for (even a part of) structural models of dynamic oligopoly. For a dynamic structural model of the representative agent, the Euler-equation-based estimation technique is usually employed (see Hansen and Singleton (1982)). However, the Euler equations cannot be generally obtained in dynamic oligopoly.[7] As an alternative, we may consider an estimation procedure in dynamic oligopoly as follows. Under some regularity conditions as in Ericson and Pakes (1995), a firm's optimal pricing (or quantity-setting) in dynamic oligopoly can be formulated as a continuous Markov decision problem (MDP). Then we may apply an estimation procedure similar to the nested fixed point algorithm in Rust (1994): using *ad hoc* assumptions for stochastic specification of the evolution of state variables, we may calculate each firm's value functions in equilibrium for each

[5] The airframe industry is another example in which learning-by-doing and strategic interactions are significant (see Benkard (1997)).
[6] Some examples of competing technologies are: VHS vs. Betamax in VCRs, MS-DOS vs. Machintosh operating system in personal computer operating systems, and Nintendo vs. Sega vs. Atari in home video game systems.
[7] In the presence of strategic interdependence, the envelope theorem does not hold in general, so the Euler equation can not be obtained. Refer to Appendix A for a detailed discussion.

candidate value of the parameter vector and then search for the value of the parameter vector that maximizes the (log) likelihood function or minimizes some distance. It is, however, impractical to implement this estimation procedure in the case of dynamic oligopoly. Most of all, it will result in a prohibitive computational burden. It is well known that continuous MDPs have the problem of Bellman's curse of dimensionality (see Pakes and McGuire (1996); Rust (1997)). Even with some simple discretization assumptions and a stochastic algorithm to break the curse of dimensionality, the computational burden to calculate the equilibrium value functions for just one candidate value of the parameter vector is usually huge (see Pakes and McGuire (1996); Benkard (1997)). In addition, the complexity of the estimation problem usually makes it difficult to determine the robustness of the conclusions to the *ad hoc* stochastic assumptions. Furthermore, if the stochastic process is misspecified, the estimator for the parameter vector is generally inconsistent.

The estimation procedure developed in this paper, however, enables us to semiparametrically estimate a class of structural models of dynamic oligopoly. It will be shown that, with a separable unobserved state variable (for example, an unobservable cost characteristic), first-order profit maximization conditions of dynamic oligopoly lead to the semiparametric regression model in (1.4) and (1.5). A technical difficulty of the regression model in (1.4) and (1.5) is that we can not eliminate the nonparametric part, $\phi(v_j)$, in the two-step procedure of a typical semiparametric regression model in (1.1) and (1.2) since the error term is correlated with the nonparametric part. However, we can still obtain a consistent and asymptotically normal estimator for the parameter vector θ_0 if (together with some regularity conditions we will discuss in section 2) there exist two sets of instrumental variables (IVs) satisfying both an identification condition and an orthogonality condition. Our estimation plan is as follows. We first filter the nonparametric part $\phi(v_j)$ by the first set of IVs in order to eliminate $\phi(v_j)$. For the identification of θ_0, we need the second set of IVs which is not a function of the first set of IVs. In order to construct moment conditions, it is required that the filtering error should be orthogonal to the second set of IVs (the orthogonality condition). Then based on the moment condition, we obtain an estimator, called a semiparametric instrumental variables (SIV) estimator, for θ_0 which will be proved to be consistent and asymptotically normal. The J statistic is employed to statistically test the moment conditions implied by our orthogonality condition. We provide two generic examples in which we can construct two sets of IVs both satisfying the orthogonality condition and not violating the identification condition, and then discuss an empirical example of the application of the SIV estimation procedure to estimate network effects in the U.S. home VCR market. This empirical example indicates that the SIV estimation procedure can be applied to some structural models of dynamic oligopoly in order to avoid a prohibitive computational burden of calculating each firm's value functions in equilibrium for each candidate value of the parameter vector.

The remainder of the paper is organized as follows. Section 2 discusses the identification and the orthogonality conditions and the SIV estimation procedure. Section 3 suggest a sufficient condition, called the conditional uncorrelation requirement, for the orthogonality and then discusses two generic examples. Section 4 provides an empirical example of the application of the SIV estimation procedure to estimate network effects in dynamic oligopoly. Section 5 concludes the paper. Appendix A shows that the Euler equations cannot be generally obtained in the presence of strategic interdependence. Appendix B provides the proof of consistency and asymptotic normality of the SIV estimator.

2. ESTIMATION PROCEDURE

In this section, we discuss the SIV estimation procedure. It will be shown that we can obtain a consistent and asymptotically normal estimator for the parameter vector θ_0 in the semiparametric regression model in (1.4) and (1.5) if (together with some regularity conditions) there exist two sets of IVs satisfying both an identification condition and an orthogonality condition. In section 3, we will discuss how to construct these IVs in specific examples. Let $z_{1j} \in R^m$ and $z_{2j} \in R^r$ (with $r \geq k$) denote the first and the second set of IVs, respectively. Then $E[\xi_j \mid z_{1j}, z_{2j}] = 0$, where ξ_j is the error term in the regression model of (1.4). In the semiparametric regression model of (1.4), x_j should not be a subvector of the first set of IVs, z_{1j}. Otherwise, θ_0 can not be identified.[8]

2.1. Filtering

Since the error term, ξ_j, is correlated with the nonparametric part, $\phi(v_j)$, of our regression model in (1.4) and (1.5), we cannot eliminate $\phi(v_j)$ in the two-step procedure of a typical semiparametric regression model in (1.1) and (1.2). In order to eliminate $\phi(v_j)$, we first filter $\phi(v_j)$ by the first set of IVs, z_{1j}. Filtering the nonparametric part $\phi(v_j)$ by z_{1j}, we can decompose $\phi(v_j)$ into its projection onto the space spanned by z_{1j}, $E[\phi(v_j) \mid z_{1j}]$, and a projection error, $\varepsilon_j \in R$, orthogonal to z_{1j}. Then by the property of conditional expectations, we have: $\varphi(z_{1j}) = E[\phi(v_j) \mid z_{1j}]$, where φ is a function from R^m to R. Hence we have:

(2.1) $\phi(v_j) = \varphi(z_{1j}) + \varepsilon_j$,

with

(2.2) $E[\varepsilon_j \mid z_{1j}] = 0$.

Substituting (2.1) into the regression model of (1.4) yields:

[8] If x_j is a subvector of z_{1j}, then $f(x_j, \theta) = E[f(x_j;\theta) \mid z_{1j}]$ for any θ, which violates the identification

(2.3) $y_j = f(x_j;\, \theta_0) + \varphi(z_{1j}) + u_j$,

where $u_j \equiv \xi_j + \varepsilon_j$. The CMRs in both (1.5) and (2.2) imply:

(2.4) $E[\, u_j \mid z_{1j}\,] = E[\, \xi_j \mid z_{1j}\,] + E[\, \varepsilon_j \mid z_{1j}\,] = 0.$

Now the CMR of (2.4) is applied to eliminate the unknown projection $\varphi(z_{1j})$ in a two-step procedure, leading to an alternative regression model:

(2.5) $y_j - E[y_j | z_{1j}] = f(x_j;\theta_0) - E[f(x_j;\theta_0) | z_{1j}] + u_j$.

By the property of conditional expectations, we let $\tau_{10}(z_{1j}) = E[y_j | z_{1j}]$ and $\tau_{20}(z_{1j},\theta) = E[f(x_j;\theta) | z_{1j}]$, where $\theta \in R^k$, τ_{10} is a function from R^m to R, and τ_{20} is a function from $R^m \times R^k$ to R. Both $\tau_{10}(z_{1j})$ and $\tau_{20}(z_{1j},\, \theta)$ can be nonparametrically estimated. Let $\tau_0(z_{1j},\theta) = (\tau_{10}(z_{1j}), \tau_{20}(z_{1j},\theta))'$.

2.2. Identification and Orthogonality

The fundamental necessary condition for the identification of θ_0 can be stated as follows.

ASSUMPTION 1 (Identification): The two vectors of IVs, z_{1j} and z_{2j}, are such that for any $\theta \neq \theta_0$,

$E[f(x_j;\theta) - \tau_{20}(z_{1j},\theta) | z_{2j}] \neq 0$ and $E[y_j - \tau_{10}(z_{1j}) - f(x_j;\theta) + \tau_{20}(z_{1j},\theta) | z_{2j}] \neq 0$.

---

condition of Assumption 1 that follows.

The fundamental necessary identification condition is required for the existence of a consistent estimator of θ_0. The CMR of (2.4) cannot be applied to the alternative regression model of (2.5) since the identification condition of Assumption 1 will be violated. In general, if z_{2j} is a function of z_{1j}, then $E[f(x_j;\theta)|z_{2j}] = E[E[f(x_j;\theta)|z_{1j}]|z_{2j}]$ for any θ.[9] Therefore, if z_{2j} is a function of z_{1j}, the fundamental necessary condition for the identification of Assumption 1 is violated.

The orthogonality condition requires that the filtering error, ε_j, should be orthogonal to the second set of IVs.

ASSUMPTION 2 (orthogonality): The two vectors of IVs, z_{1j} and z_{2j}, are such that $E[\, z_{2j}\varepsilon_j\,] = 0$.

Assumption 2 will turn out to be a sufficient condition to construct moment conditions in our SIV estimation procedure.

2.3. Two-Step Estimation Procedure

We now discuss a two-step estimation procedure based on the alternative regression model in (2.5). Le m_j: $R^k \times R^2 \rightarrow R^r$ defined as:[10]

$$(2.6) \qquad m_j(\theta,\tau_0) = z_{2j}\{y_j - \tau_{10} - f(x_j;\theta) + \tau_{20}\}\ ,$$

[9] In addition, $E[y_j|z_{2j}] = E[E[y_j|z_{1j}]|z_{2j}]$ and hence $E[y_j - \tau_{10}(z_{1j}) - f(x_j;\theta) + \tau_{20}(z_{1j},\theta)|z_{2j}] = 0$ for any θ.

[10] We can construct a GMM weight matrix, say $\gamma(z_{2j})$, in a similar fashion of Hansen and Singleton (1982). As discussed in Chamberlain (1992), however, this GMM estimator generally does not obtain the efficiency bound in the semiparametric regression cases.

where $\tau_0 = (\tau_{10}, \tau_{20})'$, $\tau_{10} = \tau_{10}(z_{1j})$ and $\tau_{20} = \tau_{20}(z_{1j}, \theta)$. Assumption 2 is a sufficient condition in which we have: $E[m_j(\theta_0, \tau_0)] = 0$.[11]

LEMMA 1: Under Assumption 2, we have: $E[m_j(\theta_0, \tau_0)] = 0$.

Proof. Since $u_j = \xi_j + \varepsilon_j$, $E[z_{2j}u_j] = E[z_{2j}\xi_j] + E[z_{2j}\varepsilon_j]$. Using the CMR that $E[\xi_j \mid z_{1j}, z_{2j}] = 0$, we have: $E[z_{2j}\xi_j \mid z_{1j}] = E[E[z_{2j}\xi_j \mid z_{1j}, z_{2j}] \mid z_{1j}] = E[z_{2j}E[\xi_j \mid z_{1j}, z_{2j}] \mid z_{1j}] = 0$.
Hence, $E[z_{2j}\xi_j] = E[E[z_{2j}\xi_j \mid z_{1j}]] = 0$. From Assumption 2, we have $E[z_{2j}\varepsilon_j] = 0$. Therefore,
$E[m_j(\theta_0, \tau_0)] = E[z_{2j}u_j] = E[z_{2j}\xi_j] + E[z_{2j}\varepsilon_j] = 0$. Q.E.D.

Based on the alternative regression model of (2.5) and the moment condition in Lemma 1, we now apply a two-step estimation procedure to obtain a SIV estimator for θ_0. In the first step, we obtain nonparametric estimators for $\tau_{10}(z_{1j})$ and $\tau_{20}(z_{1j}, \theta)$. Let $\hat{\tau}(z_{1j}, \theta)' = (\hat{\tau}_1(z_{1j}), \hat{\tau}_2(z_{1j}, \theta))$ denote the preliminary nonparametric estimator for $\tau_0(z_{1j}, \theta)' = (\tau_{10}(z_{1j}), \tau_{20}(z_{1j}, \theta))$. For notational simplicity, define: $G_J(\theta) = [\sum_{j=1}^{J} m_j(\theta, \hat{\tau})]/J$ and $G(\theta) = E[m_j(\theta, \tau_0)]$. Lemma 1 guarantees that $G(\theta_0) = 0$. We choose as a SIV estimator, say θ_J, the value which satisfies

(2.7) $\| G_J(\theta_J) \| = \inf_{\theta \in \Theta} \| G_J(\theta) \| + o_p(1/\sqrt{J})$,[12]

[11] The expectation operator $E[\cdot]$ is with respect to a distribution for $W_j = (y_j, x_j', z_{1j}', z_{2j}')'$.
[12] We let $\|A\|$ denote the Euclidean norm of a vector or matrix A, i.e., $\|A\| = (\mathrm{trace}(A'A))^{1/2}$.

where $\theta_0 \in int\Theta$, a bounded subset of R^k. Under some regularity conditions, we can prove that the SIV estimator, θ_J, defined in (2.7) is consistent and asymptotically normal.

THEOREM 1: Suppose: (i) $f(x_j;\ \theta)$ is once continuously differentiable in θ and has a square integrable envelope;[13] (ii) z_{1j} and z_{2j} satisfy the orthogonality condition in Assumption 2; (iii) for any δ, $\inf_{\|\theta-\theta_0\|>\delta} \|G(\theta)\| > 0$;[14] (iv) $\|z_{2j}\| < \infty$; (v) Let J denote the number of observations. For α > ¼, $\sup_{(z_{1j})} J^{\alpha} \|\hat{\tau}_1(z_{1j}) - \tau_{10}(z_{1j})\| = O_p(1)$, and $\sup_{(z_{1j},\theta)} J^{\alpha} \|\hat{\tau}_2(z_{1j},\theta) - \tau_{20}(z_{1j},\theta)\| = O_p(1)$;[15] (vi) Γ′Γ has full rank, where $\Gamma \equiv \partial G(\theta_0) / \partial \theta'$; and (vii) $E[y_j - \tau_{10}(z_{1j})]^2 < \infty$, and $E[f(x_j;\theta_0) - \tau_{20}(z_{1j},\theta_0)]^2 < \infty$. Then the SIV estimator θ_J defined in (2.7) is consistent and asymptotically normal with the covariance matrix given by $\Lambda = (\Gamma'\Gamma)^{-1}\Gamma' V \Gamma(\Gamma'\Gamma)^{-1}$, where $V = \lim_{J\to\infty} J(E[G_J(\theta_0)])(E[G_J(\theta_0)])'$.

The Appendix B proves Theorem 1.

3. GENERIC EXAMPLES

In general, the projection error ε_j may or may not be orthogonal to z_{2j} unless z_{2j} is a function of z_{1j}. The identification condition of Assumption 1, however, will be violated if z_{2j} is a function of z_{1j}. Hereafter z_{1j} and z_{2j} are said to be *different* sets of IVs if z_{2j} is not a function of z_{1j}.

---

[13] $f(x_j;\ \theta)$ has a square integrable envelope if and only if there is $\bar{f}(x_j)$ such that $|f(x_j;\ \theta)| \le \bar{f}(x_j)$ and $\int \bar{f}(x_j)^2 P(dx_j) \le K < \infty$, where $P(dx_j)$ is the distribution function of x_j.

[14] This is the identification condition. The condition in Assumption 1 is a necessary condition for this.

The critical question on the SIV estimation procedure may be how to construct two different sets of IVs satisfying the orthogonality condition of Assumption 2. In this section, we first provide a sufficient condition in which two different sets of IVs satisfy the orthogonality of Assumption 2, and then discuss two generic examples.

A sufficient condition in which two different sets of IVs satisfy the orthogonality will be: $Cov[\, z_{2j}, \phi(v_j) \mid z_{1j}] = 0$, which we will call *conditional uncorrelation requirement*. The conditional uncorrelation requirement means that for the given first set of IVs z_{1j}, the second set of IVs z_{2j} and the nonparametric part $\phi(v_j)$ must be linearly uncorrelated. Hence this requirement can instruct how to construct two different sets of IVs in a specific economic model, depending on the meaning of the nonparametric part and available exogenous variables. Section 4 provides such an empirical example.

LEMMA 2: If $Cov[\, z_{2j}, \phi(v_j) \mid z_{1j}] = 0$, then $E[z_{2j}\varepsilon_j\,] = 0$.

Proof: From the filtering equation in (2.1), we have:

$$E[z_{2j}\varepsilon_j \mid z_{1j}] = E[z_{2j}\phi(v_j) \mid z_{1j}\,] - E[z_{2j}\, E[\,\phi(v_j) \mid z_{1j}\,] \mid z_{1j}\,]$$

$$= E[z_{2j}\phi(v_j) \mid z_{1j}\,] - E[z_{2j} \mid z_{1j}\,]\, E[\,\phi(v_j) \mid z_{1j}\,] = Cov[z_{2j}, \phi(v_j) \mid z_{1j}].$$

Therefore, if $Cov[z_{2j}, \phi(v_j) \mid z_{1j}] = 0$, $E[z_{2j}\varepsilon_j\,] = E[\,E[z_{2j}\varepsilon_j \mid z_{1j}]\,] = 0$. Q.E.D.

Now we discuss the first generic example. In the regression model of (1.4) and (1.5), we assume: (i) x_j is a vector of exogenous variables and is excluded from the nonparametric part; and (ii) $v_j = (w_j', d_j')'$ where is $w_j \in R^w$ ($w < l$) is a vector of exogenous variables, and $d_j \in R^{l\text{-}w}$ is a vector of variables correlated with ξ_j. d_j may include ξ_j. Then Generic Form I can be written as:

[15] Some bias-reducing multivariate normal-based kernel estimators as in Bierens (1987) can be used as such preliminary estimators.

(3.1) $y_j = f(x_j; \theta_0) + \phi(w_j, d_j) + \xi_j$,

with

(3.2) $E[\, \xi_j \mid x_j, w_j\,] = 0$.

If $(x_j', w_j')'$ and d_j are stochastically independent, we can construct two different sets of IVs satisfying the conditional uncorrelation requirement as follows. We first choose $z_{1j} = w_j$. Then we choose $z_{2j} = H(x_j, w_j)$ for a given function H: $R^d \times R^w \rightarrow R^r$. As discussed in section 2, it is necessary for the identification condition of Assumption 1 that $H(x_j, w_j)$ should not be a function of only w_j.

THEOREM 2: In Generic Form I described in (3.1) and (3.2), if $(x_j', w_j')'$ and d_j are independent, then choosing $z_{1j} = w_j$ and $z_{2j} = H(x_j, w_j)$ for a given function H: $R^d \times R^w \rightarrow R^r$, the conditional uncorrelation requirement, $E[H(x_j, w_j)\varepsilon_j | w_j] = 0$, is satisfied.

Proof: The following proof is based on 2.2.10 Theorem in Florens, Mouchart and Rolin (1990). "⇒" means "imply" in this proof. Suppose that A, B, and C are random variables defined on the same sample space. If A and B are conditionally (on C) independent, we will write "A ⊥ B | C". The independence between $(x_j', w_j')'$ and d_j implies:

$$(x_j', w_j')' \perp d_j | w_j \Rightarrow (x_j', w_j')' \perp (w_j', d_j')' | w_j \Rightarrow H(x_j, w_j) \perp (\phi(w_j, d_j), w_j) | w_j$$

for any functions ϕ and H, which implies once again:

$$H(x_j, w_j) \perp (\phi(w_j, d_j) - E[\phi(w_j, d_j) | w_j]) | w_j \Rightarrow E[H(x_j, w_j)\varepsilon_j | w_j] = 0,$$

where $\varepsilon_j = \phi(w_j, d_j) - E[\phi(w_j, d_j) | w_j]$. Q.E.D.

The second generic example is a case in which a vector of *excluded* exogenous variables is available while the parametric function includes some endogenous variables. More specifically, in the regression model of (1.4) and (1.5), we assume: (i) x_j includes a vector of endogenous variables; (ii) there exists a vector of excluded (from the nonparametric part) exogenous variables, say $g_j \in R^s$;[16] and (iii) $v_j = (w_j', d_j')'$ where is $w_j \in R^w$ ($w < l$) is a vector of (included) exogenous variables, and $d_j \in R^{l-w}$ is a vector of variables correlated with ξ_j. d_j may include ξ_j. Hence Generic Form II can be written as:

(3.3) $y_j = f(x_j; \theta_0) + \phi(w_j, d_j) + \xi_j$,

with

(3.4) $E[\xi_j \mid g_j, w_j] = 0$.

Then choose $z_{1j} = w_j$ and $z_{2j} = H_g(g_j, w_j)$ for a given function H_g: $R^s \times R^w \rightarrow R^r$. As discussed in section 2, it is necessary for the identification condition of Assumption 1 that $H_g(g_j, w_j)$ should not a function of only w_j.

THEOREM 3: Consider the Generic Form II described in (3.4) and (3.5). If $(g_j', w_j')'$ and d_j are independent, then choosing $z_{1j} = w_j$ and $z_{2j} = H_g(g_j, w_j)$ for a given function H_g: $R^s \times R^w \rightarrow R^r$, the conditional uncorrelation requirement, $E[H_g(g_j, w_j)\varepsilon_j | w_j] = 0$, is satisfied.

The proof for Theorem 3 is similar to that of Theorem 2.

---

[16] In the case that f is a linear function, the identification condition requires that g_j should be correlated either with x_j or with w_j.

Possible applications of Theorem 3 may include a semiparametric estimation of structural models of dynamic oligopoly. In dynamic oligopoly, Markov perfect equilibrium is usually adopted as a solution concept. In a Markov perfect equilibrium, some exogenous state variables may have an arbitrarily small influence on the stream of future profits (see Fudenberg and Tirole (1992)). In this case, the first order conditions derived from dynamic oligopoly can be treated as an example of Generic Form II. We will detail this idea in an empirical example in section 4.

The additive semiparametric model in Newey, Powell and Vella (1999) can be considered a more general case than Generic Form II in the sense that it allows the influence of excluded exogenous variables on the nonparametic part through endogenous variables. However, the estimation procedure suggested in Newey, Powell and Vella (1999) may not be practical in structural estimation of oligopoly. In structural estimation of oligopoly, we usually have endogenous variables (prices or quantities) both as dependent variables and as a part of explanatory variables. If the estimation procedure in Newey, Powell and Vella (1999) is applied to this case, it requires preliminarily estimating reduced-form equations of endogenous variables in order to estimate the structural-form equations of the same endogenous variables. In addition, the preliminary reduced form estimation may suffer from the curse of dimensionality or the specification problem in this case.

4. EMPIRICAL EXAMPLE

An example of the regression model in (1.4) and (1.5) can be derived from structural models of dynamic oligopoly in general. Dynamic oligopoly is a situation in which firms' price-settings (or quantity-settings) are strategically interdependent and have durable effects on the stream of their future profits. Dynamic oligopoly fits many industries characterized by the presence of network externalities, informational product differentiation, or learning-by-doing. In the presence of network externalities or informational product differentiation, a firm's current output level affects its installed base or consumers' experience utilities and thus its demands and profits in the future. In the presence of learning-by-doing, a firm's current output level affects its future marginal production cost and thus profits. In this section, we discuss an empirical example of applying the SIV estimation procedure to estimate network effects in the framework of Park (1999).

In dynamic oligopoly, a firm makes a dynamic optimization decision, based on available information, beliefs on the rivals' strategies and beliefs on the evolution of the state. Usually dynamic structural models assume rational expectations and Markov processes for the evolution of the state variables. Let $\Omega_t = (X_t', \xi_t', \omega_t')'$ denote a vector of exogenous state variables in period t, where X_t is a vector of observable product characteristics, and $(\xi_t', \omega_t')'$ is a vector of unobservable product and cost characteristics. In many applied microeconomics studies, the existence and importance of the unobservable state variable have been recognized (see Rust (1994); Bery, Levinsohn, and Pakes (1995); Park (1999)). Let B_t denote a vector of cumulative output levels (endogenous state variables), and p_t denote a vector of price levels (continuous decision variables) in period t. Let subscript j indicate a vector of firm j's variables, and subscript $-j$ indicate a vector of variables of firm j's rivals. Hence, for example, $\Omega_t' = (\Omega_{jt}', \Omega_{-jt}')$. Under some regularity conditions,[17] firm j's optimal decision rule solves for the Bellman equation:

$$(4.1) \qquad v(\Omega_{jt}, \Omega_{-jt}, B_{jt}, B_{-jt}) = \sup_{p_{jt}} \{ \Pi_j(\Omega_t, B_t, p_t; \theta_0) +$$

$$\beta \int v(\Omega_{jt+1}, \Omega_{-jt+1}, B_{jt+1}, B_{-jt+1}) \, \mathrm{P}_B(dB_{t+1}|B_t, p_t) \, \mathrm{P}_\Omega(d\Omega_{t+1}|\Omega_t)\},$$

[17] For the regularity conditions, refer to Ericson and Pakes (1995) and the literature cited there.

where $\Pi_j(\cdot)$ is a single-period payoff function, β is the discount factor, $P_\Omega(\cdot|\cdot)$ is the conditional distribution of the next-period exogenous state vector, Ω_{t+1}, given the current-period exogenous state vector, Ω_t, and $P_B(\cdot|\cdot)$ is the conditional distribution of the next period endogenous state vector, B_{t+1}, given the current period endogenous state vector, B_t, and the current decision vector, p_t.

Then, the dynamic first-order conditions of (4.1) can be derived as follows:[18]

$$(4.2) \quad \frac{\partial \Pi_j(\Omega_t, B_t, p_t)}{\partial p_{jt}} + \beta \int v(\Omega_{jt+1}, \Omega_{-jt+1}, B_{jt+1}, B_{-jt+1}) \frac{\partial P_B(dB_{t+1}|B_t, p_t)}{\partial p_{jt}} P_\Omega(d\Omega_{t+1}|\Omega_t) = 0.$$

Ericson and Pakes (1995) provided the regularity conditions in which the dynamic optimization behavior based on (4.2) generates a Markov perfect equilibrium. The second term on the left-hand side of (4.2) indicates that the firm takes into account the effects of current action (e.g., pricing) on its own and its rivals' future endogenous state variables (e.g., cumulative outputs) and thus its future profit stream. Due to the effects of a firm's current action on its rivals' future endogenous state variables, the envelope theorem does not hold, and thus the Euler equations can not be generally obtained. Refer to Appendix A for a formal proof for this. A theoretical alternative to the Euler-equations-based estimation procedure can be as follows. Using *ad hoc* assumptions for stochastic specification of the evolution of state variables, we may calculate each firm's value functions in equilibrium for each candidate value of the parameter vector and then search for the value of the parameter vector that maximizes the (log) likelihood function or minimizes some distance. This alternative procedure, however, will cause a prohibitive computational burden. Even with some simple discretization assumptions and a stochastic algorithm, the computational burden to calculate the equilibrium value functions for just one candidate value of the parameter vector is usually huge (see Pakes and McGuire (1996); Benkard (1997)).

Instead, we may derive a semiparametric regression model from the dynamic first order conditions in (4.2). Suppose that the marginal cost function, mc_j, of product j is the sum of a hedonic function, say $\Gamma(X_j)$, and an unobserved cost characteristic, ω_j: $mc_j = \Gamma(X_j) + \omega_j$. Then the dynamic first-order condition in (4.2) leads to the following pricing equation of dynamic oligopoly:

(4.3) $$p_{jt} = -\frac{1}{\partial q_{jt} / \partial p_{jt}} q_{jt} + \phi_{jt} + \omega_{jt},$$

where $\phi_{jt} = \phi_j(B_t, \Omega_t) =$

$$\Gamma(X_{jt}) + \frac{-1}{\partial q_{jt} / \partial p_{jt}} \beta \int V(\Omega_{jt+1}, \Omega_{-jt+1}, B_{jt+1}, B_{-jt+1}) \frac{\partial \mathrm{P}_B(dB_{t+1} | B_t, p_t)}{\partial p_{jt}} \mathrm{P}_\Omega(d\Omega_{t+1} | \Omega_t).$$

Note that prices and sales can be expressed as functions of state variables. The second term of ϕ_{jt} in (4.3) is a mark-up reflecting the effects of the firm's current price on its own and its rivals' future cumulative outputs and thus the stream of its future profits. For a given (parametric) demand function, $-q_{jt} / (\partial q_{jt} / \partial p_{jt})$ in (4.3) is a known function with a parameter vector. However, since ϕ_{jt} (especially its second term) is an unknown function, the pricing equation of (4.3) is of the form of a semiparametric regression model. Indeed, this equation is an example of our semiparametric regression model in (1.3) and (1.4) since the error term ω_{jt} is correlated with the nonparametric part ϕ_{jt}. Note that ω_{jt} is a part of the state vector Ω_t. In addition, it is correlated with current prices and sales in oligopoly, which may affect the producer's future profits.

---

[18] We assume that $\Pi_j(\cdot)$ and $P_B(\cdot|\cdot)$ are differentiable with respect to p_{jt}.

Now we discuss the framework developed in Park (1999) to estimate network effects in dynamic oligopoly. Network externalities may be understood as positive consumption externalities: in the presence of network externalities, an increase in the number of users of a product raises the consumer's utility level and hence the demands for that product. The number of users is usually called a network size, and the user's benefit from the network size is called a network benefit. In many examples of network externalities, products such as VCRs and the computer operating systems are durable goods. In the case of durable goods, the consumer may take into account not only the current utility but also the expected future utilities derived from the use of a product. This dynamic concern of the consumer is represented by the consumer's value function in Park (1999). We consider a situation in which consumers choose a format (VHS or Betamax in the VCR case) first and then a brand within the format. If two products are of the same format, then they are compatible. We assume that products are differentiated. Then in Park (1999), the consumer's value function is derived as the sum of the (average) network benefit and the (average) stand-alone benefit of the product. More specifically consumer *i*'s (average) value function for product *j* of format *g* in period *t*, say $V_{ij,t}$, is defined as:

$$(4.4) \qquad V_{ij,t} = N_{gt} - \varphi\lambda p_{jt} + X_j\alpha + \xi_j + \zeta_{ig} + (1-\sigma)\varepsilon_{ij} \,,$$

and

$$N_{gt} = \varphi \sum_{s \geq t} (1-\varphi)^{s-t} E[NB_{gs} | \Omega_t, B_t],$$

where NB_{gt} is a single-period network benefit of format *g* in period *t*, $\zeta_{ig} + (1\text{-}\sigma)\varepsilon_{ij}$ is consumer *i*'s idiosyncratic taste for differentiated product *j* of format *g*, φ is the consumer's discount rate, and $\theta_0{'} = (\varphi\lambda, \sigma, \alpha', \{N_{gt}\}_t)$ is the parameter vector which we want to estimate. N_{gt} represents the (average) network benefit from using product *j* from period *t* on, and

$-\varphi\lambda p_{jt} + X_j a + \xi_j + \zeta_{ig} + (1-\sigma)\varepsilon_{ij}$ represents consumer *i*'s (average) utility level for the attributes of product *j*. Using the nested logistic assumptions as in Cardell (1997), closed-form demand (or market share) functions can be derived from the consumer's value function in (4.4).

From the obtained market share functions, we derive a demand-side estimating equation. The obtained market share functions also imply that the parametric part of the pricing equation in (4.3) will be: $-q_{jt}/(\partial q_{jt}/\partial p_{jt}) = (1-\sigma)/\{(1-\sigma S_{j/gt} - (1-\sigma)S_{jt})\varphi\lambda\}$, where S_{jt} and $S_{j/gt}$ denote product *j*'s market share and product *j*'s within-format market share, respectively. Then we have a system of the demand-side and producer-side estimating equations as follows:

$$\text{(4.5)} \quad \begin{bmatrix} \ln(S_j / S_0) \\ p_j \end{bmatrix} = \begin{bmatrix} -c - \varphi\lambda p_j + X_j\alpha + \sigma \ln S_{j/g} + N_g + \xi_j \\ (1-\sigma)/([1-\sigma S_{j/g} - (1-\sigma)S_j]\varphi\lambda) + \phi_j + \omega_j \end{bmatrix}.$$

Henceforth, it is understood that variables are indexed by time. Our main interest is to estimate the network effects N_g. Since N_g is common to all products of format *g*, N_g and σ cannot be identified by using the demand-side estimating equation alone. For the identification of N_g and σ, we need the pricing equation.[19]

As discussed above in (4.3), the producer-side estimating equation is an example of our generic regression model in (1.4) and (1.5). Hence we need to construct two different sets of instrumental variables satisfying the conditional uncorrelation requirement. Since prices are determined endogenously in oligopoly, we also need a vector of instrumental variables for the demand-side estimating equation. We assume the stochastic independence between observable product characteristics X_t and unobservable characteristics $(\xi_t', \omega_t')'$ and installed bases B_t:

ASSUMPTION 3: The $(X_t', \xi_t', \omega_t', B_t)'$ vectors have the property that X_t and $(\xi_t', \omega_t', B_t')'$ are stochastically independent.

Then as in Berry, Levinsohn and Pakes (1995), $(X_j', X_{-j}')'$ can be employed as instrumental variables for the demand-side estimating equation since a firm's price is correlated with its own and its rivals' observed product characteristics. The observed product characteristics in the VCR data of Park (1999) include high quality (HQ), number of programmable events (events), on-screen display (osd), multichannel TV sound decoder (mts), stereo, and hi-fi.[20] The unobserved characteristic reflects the heterogeneity of VCR producers in marketing ability, reliability of the brand, etc.

In preliminary estimations of demand-side equation alone,[21] we have found that observable product characteristics are not statistically significant in our case. Hence we suspect that these variables can be candidates for the exogenous state variables which may have an arbitrarily small influence on the stream of future profits. More specifically, we expect that for the given observable product characteristics of a firm, the observable product characteristics of the rivals (both in the same format and in the other format) may be correlated neither with the firm's hedonic marginal cost nor with the stream of the firm's future profits (or mark-up). Hence the observable product characteristics of the rivals, X_{-j}, are treated as excluded exogenous variables in our case. The included exogenous variables are its own observable product characteristics, X_j, in our case. The other variables, say d_j as in Generic Form II, appearing in the nonparametric part ϕ_j include a vector of endogenous state variables $(B_g, B_{-g})'$ and a vector of unobservable characteristics $(\xi_{-j}', \omega_{-j}', \xi_j, \omega_j)'$ in our case. The independence assumption in

[19] Refer to Berry (1994) for a more detailed discussion of this identification problem.

[20] "HQ" means that the VCR provides improved picture quality; "events" indicates the number of programs the VCR can be set to record automatically; "osd" is a feature that makes it possible to use on-screen commands for entering time, date, and channel of the program the user wants to tape; "mts" allows stereo TV programs to be taped or heard in stereo; "stereo" and "hifi" are features of sound quality.

[21] For example, fixing σ to be 0.8 (or other values between 0.7 and 0.95), we estimate the demand-side equation in (4.5) alone.

Assumption 3 guarantees the independence between $(X_{-j}, X_j)'$ and d_j. Hence we can select $z_{1j} = X_j$ and $z_{2j} = X_{-j}$ for the producer-side estimating equation as instructed in Theorem 3.

Table 1 reports the SIV estimates based on these IVs.[22] We use an unbalanced firm-level panel data for the U.S. home VCR market from year 1981 to year 1988 except year 1985.[23] J statistic in our calculation is 17.79 with degrees of freedom 12. Hence the over-identification restriction is accepted with the significance level of 0.1, which implies the moment condition constructed by the two sets of IVs described above is statistically valid. The within-format correlation coefficient, σ, is significant. The estimate of the within-format correlation coefficient, σ, is 0.805, which implies that VCRs of the same format are considered as roughly homogeneous products. This estimate, however, is not as high as we expected, for example, 0.9 or higher since there was almost no difference among the different brand VCRs of the same format in performance or features.[24] Prices are in \$10,000 units and deflated by the Consumer Price Index. The parameter of 'price' is significant with the significance level of 0.07 but has a very small estimated value. All product characteristics turn out to be insignificant as in the preliminary estimations. Based on the very low value of the price parameter and the insignificant estimates of product characteristics, we can infer that the differences among the brands within a format are caused by the heterogeneity of VCR producers in marketing ability, reliability of the brand, etc., which is reflected in the unobserved product characteristic of our structural regression model. All

[22] In the actual calculation, we use matrices of first-order polynomial basis functions of instrumental variables. Pakes (1994) showed that the dimension of the basis for polynomials of a given order is independent of the number of products, if the equilibrium is "partially exchangeable" (that is, exchangeable in the state vectors of a product's competitors). We assume the following two forms of exchangeability: (i) exchangeable in the order of the competing formats, and (ii) for a given format, exchangeable in the order of competing products. In our case, there are two formats, VHS and Betamax, and 6 observed product characteristics. In addition, we add the number of producers in each format as IVs. Hence we have 20 demand-side IVs and 14 producer-side second-set IVs.

[23] Refer to Park (1999) for the detailed description of the data set.

[24] As argued in Berry (1994), the within-format correlation coefficient approaches 1 as the within-format correlation of utility levels goes to 1.

the dummy variables for network effect and a constant term in each year ($N_g - c$ in (4.5)) are significant.

Park (1999) showed that the logarithm of the relative sales (log relative sales, hereafter) of the two formats can be decomposed into the sum of two differences as follows:

$$\text{(4.6)} \qquad \ln(q_v / q_b) = (1-\sigma)[\ln(\textstyle\sum_{j\in J_v} e^{\delta_j/(1-\sigma)}) - \ln(\textstyle\sum_{j\in J_b} e^{\delta_j/(1-\sigma)}] + [N_v - N_b],$$

where v denote VHS and b denote Betamax. The first term of the right-hand side in (4.6), if it is positive, represents the price/quality advantage of VHS over Betamax, while the second term stands for the network advantage of VHS over Betamax. Using the estimates of $N_g - c$ for each format in each year, we can calculate the ratio of network advantage of the VHS format over the Betamax format ($N_v - N_b$) to log relative sales of the VHS format to the Betamax format ($\log(q_v / q_b)$) in each year as reported in Table 2. Table 2 indicates that the network advantage of VHS explains 70.3% to 86.8% of the log relative sales of VHS to Betamax in each year. These numbers indicate: during the years of 1981-1988, the network advantage of VHS was the key reason that VHS outsold Betamax in the U.S. home VCR market. Park (1999) proceeded further to conclude that the network advantages of VHS over Betamax were due to VHS's larger installed base, average price advantage, bigger lineups induced by entry and exit.

5. CONCLUDING REMAKS

This paper posits a semiparametric regression model in which the error term is correlated with the nonparametric part and discusses conditions to obtain a semiparametric estimator, called SIV estimator, with consistency and asymptotic normality. Then the paper provides two generic examples and an empirical example of the application of the SIV estimation procedure. The empirical example shows that the SIV estimation procedure can be applied to some structural models of dynamic oligopoly in order to avoid a prohibitive computational burden of calculating each firm's value functions in equilibrium for each candidate value of the parameter vector.

APPENDIX A

PROPOSITION: The Euler equations cannot be generally obtained in the presence of strategic interdependence.

Proof: We will consider the following simple, deterministic case. The proof for a general, stochastic case is similar. Suppose that there are two firms, 1 and 2, which are strategically interdependent. Let (y_1, y_2) be a vector of firm 1's and firm 2's actions, and (x_1, x_2) be a vector of firm 1's and firm 2's state variables. As a simplest case, assume that y_i will be firm i's state variable in the next period. Then under some regularity conditions, firm 1's value function solves for the Bellman equation:

$$v(x_1, x_2) = \max_{y_1} [F(x_1, y_1, y_2) + \beta v(y_1, y_2)].$$

Let $g_1(x_1, x_2)$ and $g_2(x_1, x_2)$ be the optimal decision rules for firm 1 and firm 2, respectively. For notational simplicity, let F_i and v_i denote the partial derivatives of $F(\cdot)$ and $v(\cdot)$ with respect to the i^{th} element, respectively.

Then, the first order conditions are:

$$\text{(A.1)} \quad F_2[x_1, g_1(x_1, x_2), g_2(x_1, x_2)] + \beta v_1[g_1(x_1, x_2), g_2(x_1, x_2)] = 0.$$

The envelope condition is *not* satisfied because

(A.2) $$v_1(x_1,x_2) = F_1[x_1, g_1(x_1,x_2), g_2(x_1,x_2)] +$$

$$F_2[x_1, g_1(x_1,x_2), g_2(x_1,x_2)]\frac{\partial g_1(x_1,x_2)}{\partial x_1} + F_3[x_1, g_1(x_1,x_2), g_2(x_1,x_2)]\frac{\partial g_2(x_1,x_2)}{\partial x_1}$$

$$+\beta\{v_1[g_1(x_1,x_2), g_2(x_1,x_2)]\frac{\partial g_1(x_1,x_2)}{\partial x_1} + v_2[g_1(x_1,x_2), g_2(x_1,x_2)]\frac{\partial g_2(x_1,x_2)}{\partial x_1}\}$$

$$= F_1[x_1, g_1(x_1,x_2), g_2(x_1,x_2)] +$$

$$\{F_3[x_1, g_1(x_1,x_2), g_2(x_1,x_2)] + \beta v_2[g_1(x_1,x_2), g_2(x_1,x_2)]\}\frac{\partial g_2(x_1,x_2)}{\partial x_1}.$$

Note that $\partial g_1(x_1, x_2) / \partial x_1 = 0$. Now set $x_1 = x_{1t}$, $x_2 = x_{2t}$, $g_1(x_1, x_2) = g_1(x_{1t}, x_{2t}) = x_{1t+1}$ and $g_2(x_1, x_2) = g_2(x_{1t}, x_{2t}) = x_{2t+1}$ in (A.1) to obtain

$$F_2(x_{1t}, x_{1t+1}, x_{2t+1}) + \beta v_1(x_{1t+1}, x_{2t+1}) = 0,$$

and set $x_1 = x_{1t+1}$, $x_2 = x_{2t+1}$, $g_1(x_1, x_2) = g_1(x_{1t+1}, x_{2t+2}) = x_{1t+2}$ and $g_2(x_1, x_2) = g_2(x_{1t+1}, x_{2t+1}) = x_{2t+2}$ in (A.2) to obtain

$$v_1(x_{1t+1}, x_{2t+1}) = F_1[x_{1t+1}, x_{1t+2}, x_{2t+2}] + \{F_3[x_{1t+1}, x_{1t+2}, x_{2t+2}] + \beta v_2(x_{1t+2}, x_{2t+2})\}\frac{\partial x_{2t+2}}{\partial x_{1t+1}}.$$

Eliminating $v_1(x_{1t+1}, x_{2t+1})$ between these two equations then gives:

(A.3) $$F_2(x_{1t}, x_{1t+1}, x_{2t+1}) + \beta F_1[x_{1t+1}, x_{1t+2}, x_{2t+2}] +$$

$$\beta\{F_3[x_{1t+1}, x_{1t+2}, x_{2t+2}] + \beta v_2(x_{1t+2}, x_{2t+2})\}\frac{\partial x_{2t+2}}{\partial x_{1t+1}} = 0\,.$$

Due to the interaction term, (A.3) is not the Euler equation. Q.E.D.

APPENDIX B

Let $W_j' = (y_j, x_j', z_{1j}', z_{2j}')$. Suppose that $\{W_j\}_{j = 1, 2, \ldots}$ is a random draw. For technical reasons, we first redefine $m_j(\cdot)$ of (2.6) over Θ, a bounded subset of R^k, as follows:

$$m_j(\theta, \hat{\tau}) = \zeta(W_j) z_{2j}\{y_j - \hat{\tau}_1(\mathrm{z}_{1j}) - f(\mathrm{x}_j; \theta) + \hat{\tau}_2(\mathrm{z}_{1j}, \theta)\}\,,$$

where $\zeta(W_j) =$ 1 if $W_j \in W^*$ and = 0 otherwise, and W^* is an open bounded set with a minimally smooth boundary.[25] For notational simplicity, let $m_j(\theta) = m_j(\theta, \tau_0)$. Then

$$\text{(B.1)} \quad m_j(\theta, \hat{\tau}) = m_j(\theta) - \mathrm{z}_{2j}\{\hat{\tau}_1(\mathrm{z}_{1j}) - \tau_{10}(\mathrm{z}_{1j})\} + \mathrm{z}_{2j}\{\hat{\tau}_2(\mathrm{z}_{1j}, \theta) - \tau_{20}(\mathrm{z}_{1j}, \theta)\}\,.$$

In the following proofs of consistency and asymptotic normality of the SIV estimator, θ_J, defined in (2.7), it is assumed that $\Pr(\hat{\tau} \in T) \to 1$ as $J \to \infty$, where T is a family of functions with bounded derivatives up to a certain order as required below. The following proof of Theorem 1 can be considered a special case of Pakes and Olley (1995).

---

[25] $\zeta(W_j)$ is called a trimming function. All functions in consideration are assumed to be defined on W^*, and constant elsewhere. Andrews (1995, section 4.3) discussed the tradeoff between the smoothness condition and stochastic equicontinuity conditions.

B.1. Proof of Consistency

Selecting proper bandwidths, we derive bias-reducing multivariate normal-based kernel estimators, say $\hat{\tau}_1(z_{1j})$ and $\hat{\tau}_2(z_{1j},\theta)$, satisfying the condition (v) of Theorem 1, which implies: $\sup_{(\mathrm{z}_{1j})}\|\hat{\tau}_1(\mathrm{z}_{1j})-\tau_{10}(\mathrm{z}_{1j})\|=O_p(1)$ and $\sup_{(\mathrm{z}_{1j},\theta)}\|\hat{\tau}_2(\mathrm{z}_{1j},\theta)-\tau_{20}(\mathrm{z}_{1j},\theta)\|=O_p(1)$. Together with these results, the conditions (i) and (iv) of Theorem 1 imply that $m_j(\theta,\hat{\tau})$ in (B.1) is once continuously differentiable in θ and has a square integrable envelope. Since the family $\{m_j(\theta)\}$ is Euclidean for an integrable envelope,[26] we have: $\sup_\theta\|G_J(\theta)-\mathrm{G}(\theta)\|=$

$\sup_\theta\|J^{-1}\sum_j m_j(\theta,\hat{\tau})-E[m_j(\theta)]\| \le \sup_\theta\|J^{-1}\sum_j m_j(\theta)-E[m_j(\theta)]\|$ +

$\|J^{-1}\sum_j \mathrm{z}_{2j}\{\hat{\tau}_1(\mathrm{z}_{1j})-\tau_{10}(\mathrm{z}_{1j})\}\|$ + $\sup_\theta\|J^{-1}\sum_j \mathrm{z}_{2j}\{\hat{\tau}_2(\mathrm{z}_{1j},\theta)-\tau_{20}(\mathrm{z}_{1j},\theta)\}\|$ = $o_p(1)$, which, together with the condition (iii) of Theorem 1, implies that the SIV estimator, θ_J, defined in (2.7) is a consistent estimator.[27] Q.E.D.

B.2. Asymptotic Normality

Now we discuss the conditions for asymptotic normality of the SIV estimator, θ_J, defined in (2.7). First of all, we need stochastic equicontinuity of the following empirical processes:[28]

$$v_{1\mathrm{J}}(\tau_1)=\frac{1}{\sqrt{J}}\sum_{\mathrm{j}}\{\tau_1(\mathrm{z}_{1j})-\tau_{10}(\mathrm{z}_{1j})-E[\tau_1(\mathrm{z}_{1j})-\tau_{10}(\mathrm{z}_{1j})]\}\ \text{at}\ \tau_{10},$$

---

[26] A family is *Euclidean* if it is differentiable in its index set which is bounded.
[27] For details, refer to Corollary 3.2 in Pakes and Pollard (1989).
[28] An empirical process, say $\upsilon_J(\delta_J)$, defined on a metric space is called *stochastically equicontinuous* with respect to its index set if for any two sequences of random indices, say δ_{1J} and δ_{2J}, we have $\upsilon_J(\delta_{1J})$ - $\upsilon_J(\delta_{2J})$ =

and

$$v_{2J}(\tau_2,\theta)=\frac{1}{\sqrt{J}}\sum_{j}\{\tau_2(z_{1j},\theta)-\tau_{20}(z_{1j},\theta)-E[\tau_2(z_{1j},\theta)-\tau_{20}(z_{1j},\theta)]\}\ \text{at}\ (\tau_{20},\theta_0).$$

A sufficient condition for stochastic equicontinuity of a family of functions is that the family has a square-integrable envelope and satisfies Pollard’s entropy condition (see Andrews (1994(b)). Both the existence of square-integrable envelopes and Pollard's entropy conditions for each of the families of { $\tau_1(z_{1j})-\tau_{10}(z_{1j})$ } and { $\tau_2(z_{1j},\theta)-\tau_{20}(z_{1j},\theta)$ } can be ensured by the proper choice of the bias-reducing multivariate normal based kernel estimators. More specifically, these kernel estimators should have rates of convergence greater than ¼ as specified in the condition (v) of Theorem 1. Then we have:

(B.2) $\quad \sqrt{J}G_J(\theta_0)\xrightarrow{d}N(0,V).$

A proof of (B.2) is as follows:

Proof: From (B.1), we have:

$$\sqrt{J}G_J(\theta_0)=\frac{1}{\sqrt{J}}\sum_j m_j(\theta,\hat{\tau})=$$

$$\frac{1}{\sqrt{J}}\sum_j[m_j(\theta_0)-z_{2j}\{\hat{\tau}_1(z_{1j})-\tau_{10}(z_{1j})\}+z_{2j}\{\hat{\tau}_2(z_{1j},\theta_0)-\tau_{20}(z_{1j},\theta_0)\}].$$

$o_p(1)$ whenever $\|\delta_{1J}-\delta_{2J}\|=o_p(1)$. For a good overview of stochastic equicontinuity with applications to econometrics, refer to Andrews (1994(b)).

Let $g_1(W_j) = [y_j - \tau_{10}(z_{1j})]$, and $g_2(W_j) = [f(x_j;\theta_0) - \tau_{20}(z_{1j},\theta_0)]$. Then $E[g_i(W_j)] = 0$ and $E[g_i(W_j)^2] < \infty$ [29], for i = 1 and 2. Using the stochastic equicontinuities of $v_{1J}(\tau_{10})$ and $v_{2J}(\tau_{20}, \theta_0)$, we have:

$$\frac{1}{\sqrt{J}}\sum_j\{\hat{\tau}_1(z_{1j}) - \tau_{10}(z_{1j})\} = \frac{1}{\sqrt{J}}\sum_j g_1(W_j) + o_p(1),$$

and

$$\frac{1}{\sqrt{J}}\sum_j\{\hat{\tau}_2(z_{1j},\theta_0) - \tau_{20}(z_{1j},\theta_0)\} = \frac{1}{\sqrt{J}}\sum_j g_2(W_j) + o_p(1).$$

Hence, $\sqrt{J}G_J(\theta_0) = \frac{1}{\sqrt{J}}\sum_j[m_j(\theta_0) + g_1(W_j) + g_2(W_j)] + o_p(1)$. Since the family $\{m_j(\theta_0)\}$ is Euclidean for an integrable envelope, $\sqrt{J}G_J(\theta_0)$ has an asymptotic Normal distribution by the Lindberg-Levy central limit theorem. Consequently, the covariance matrix V can be calculated as follows:

$$V = E[z_{2j}V_0 z_{2j}'],$$

where

$$V_0 = E[\{\rho_j(\theta_0) + g_1(W_j) + g_2(W_j)\}\{\rho_j(\theta_0) + g_1(W_j) + g_2(W_j)\}'|z_{2j}]$$

with $\rho_j(\theta_0) = y_j - \tau_{10}(z_{1j}) - f(x_j;\theta_0) + \tau_{20}(z_{1j},\theta_0)$. Q.E.D.

Based on the result in (B.2), we can follow the Appendix A.2 of Andrews (1994) to prove that the SIV estimator, θ_J, defined in (2.7) satisfies:

[29] Refer to condition (vii) of Theorem 1.

$$\sqrt{\mathrm{J}}(\theta_{\mathrm{J}} - \theta_0) \xrightarrow{d} N(0, (\Gamma'\Gamma)^{-1}\Gamma'\mathrm{V}\Gamma(\Gamma'\Gamma)^{-1}).$$

REFERENCES

AI, C. (1997) "A Semiparametric Maximum Likelihood Estimator," *Econometrica*, 65, 933-964.

ANDREWS, D.W.K. (1991): "Asymptotic Normality of Series Estimators for Various Nonparametric and Semiparametric Models," *Econometrica*, 59, 307-345.

_______ (1994(a)): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43-72.

_______ (1994(b)): "Empirical Process Methods in Econometrics," Chapter 2 in *Handbook of Econometrics*, Vol. 4, Now York: North Holland.

_______ (1995): "Nonparametric Kernel Estimation for Semiparametric Models," *Econometric Theory*, 11, 560 – 596.

BENKARD, C. L. (1997): "Dynamic Equilibrium in the Commercial Aircraft Market," mimeo., Yale University.

BERRY, S. (1994): Estimating Discrete Choice Models of Product Differentiation," *RAND Journal of Economics*, 25, 242 – 262.

BERRY, S., J. LEVINSOHN and A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841 - 890.

BIERENS, H. J. (1987): "Kernel Estimators of Regression Functions," in *Advances in Econometrics: Fifth World Congress*, Vol. 1, ed. by T. F. Bewley, New York: Cambridge University Press.

CURRIE, G and S. PARK (1999): "Estimating the Effects of Marketing and Consumption

Experience on Demands for Antidepressant Drugs," mimeo., SUNY at Stony Brook.

CHAMBERLAIN, G. (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567-596.

ERICSON, R. and A. PAKES (1995): "Markov Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economics Studies*, 62, 53-82.

FLORENS, J.-P., MOUCHART, M. and J.-M. ROLIN (1990): *Elements of Bayesian Statistics*, New York: Marcel Dekker, Inc.

FUDENBERG, D. and J. TIROLE (1992): *Game Theory*, Cambridge: MIT Press.

HANSEN, L. (1982): "Large Sample Properties of Method of Moments Estimators," *Econometrica*, 50, 1029-1054.

HANSEN, L. and K. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269-1286.

NEWEY, W. K. and J. L. POWELL (1989): "Instrumental Variables Estimation for Nonparametric Models," mimeo., MIT.

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

PAKES, A. (1994): "Dynamic Structural Models, Problems and Prospects: Mixed Continuous Discrete Controls and Market Interactions," in *Advances in Econometrics*, ed. by J. J. Laffont and C. Sims.

PAKES, A. and P. MCGUIRE (1996): "Stochastic Algorithm for Dynamci Models: Markov Perfect Equilibrium and the 'Curse' of Dimensionality," manuscript, Yale University.

PAKES, A. and S. OLLEY (1995): "A Limit Theorem for a Smooth Class of Semiparametric Estimators," *Journal of Econometrics*, 65, 295-332.

PAKES, A. and D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

PARK, S. (1999): "Quantitative Analysis of Network Externalities in Competing Technologies: the VCR Case," mimeo., SUNY at Stony Brook.

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

RUST, J. (1994): "Dynamic Structural Models: Problems and Prospects: Discrete Decision Processes", in *Advances in Econometrics,* ed. by J. J. Laffont and C. Sims.

_______ (1997): "Using Randomization to Break the Curse of Dimensionality," *Econometrica*, 65, 487-516.

SCHERER, F. M. (1996): *Industry, Structure, Strategy, and Public Policy*, HarperCollins College Publishers.

SCHMALENSEE, R. (1982): "Product Differentiation Advantages of Pioneering Brand," *American Economic Review*, 72, 349-365.