Supplement to "Information structure and statistical information in discrete response models": Technical report

SHAKEEB KHAN
Department of Economics, Boston College

DENIS NEKIPELOV
Department of Economics, University of Virginia

In this technical report, we derive optimal convergence rates and efficiency bounds for a class of models studied in Khan and Nekipelov (2010). These models include the triangular discrete response models and simultaneous models.

SB.1. OPTIMAL RATE FOR ESTIMATION OF THE INTERACTION PARAMETER IN A TRIANGULAR DISCRETE RESPONSE MODEL

The fact that the information associated with the treatment effect coefficient parameter, α_0, is zero (Theorem 2.1 in Khan and Nekipelov (2010)) does not imply that this parameter cannot be estimated consistently. We now provide the convergence rates of the semiparametric estimator for α_0. This will prove useful when we consider specific distributional assumptions in Section SB.3.

We begin with a definition of the optimal rate following Ibragimov and Has'minskii (1978). Let $\mathcal{G}$ characterize a class of joint densities of error terms (U, V) (denoted $g(\cdot, \cdot)$) and single indices X_1 and X (with density function $f(\cdot, \cdot)$). By $P_{f,g}$ we denote the probability measure associated with the product of two densities f and g. Suppose that $\widehat{\alpha}$ is a consistent estimator for parameter α_0. First, we recall that for the class of distributions $\mathcal{G}$, we define the risk using a positive (rate) sequence r_n and a constant L as

$$R(\hat{\alpha}, r_n, L) = \sup_{f,g \in \mathcal{G}} P_{f,g}\big(r_n|\hat{\alpha} - \alpha_0| \geq L\big).$$

Using this notion of the risk, we introduce the definition of the convergence rates for the estimator.

DEFINITION SB.1. (i) We call the positive sequence r_n the lower rate of convergence for the class of densities $\mathcal{G}$ if there exists $L > 0$ such that

$$\liminf_{n\to\infty} \inf_{\hat{\alpha}} R(\hat{\alpha}, r_n, L) \geq p_0 > 0 \quad \text{for some constant } p_0.$$

Shakeeb Khan: shakeeb.khan@bc.edu
Denis Nekipelov: denis@virginia.edu

(ii) We call the positive sequence r_n the upper rate of convergence if there exists an estimator $\hat{\alpha}_n$ such that

$$\lim_{L\to\infty} \limsup_{n\to\infty} R(\hat{\alpha}_n, r_n, L) = 0.$$

(iii) The positive sequence r_n is the minimax (or optimal) rate of convergence if it is both a lower and an upper rate.

We derive the upper convergence rate by providing an estimator that attains the upper rate of convergence in Definition SB.1(ii). The convergence rate of the estimator relies on the tail behavior of the joint density of the error distribution. We formulate assumptions that restrict the "thickness" of tails of the error distribution in addition to Assumption SB.1, which requires that the density of this distribution is smooth and the random shocks U and V are independent from the covariates X_1 and X.

ASSUMPTION SB.1. (i) *The single indices X_1 and X have a joint distribution with the full support on $\mathbb{R}^2$ that is not contained in any proper one-dimensional subspace. The parameter of interest is in the interior of a convex compact set $\mathcal{A}$.*

(ii) *The shocks (U, V) are independent of X_1 and X and have an absolutely continuous density with full support on $\mathbb{R}^2$ and joint c.d.f. $G(\cdot, \cdot)$. The partial derivative $\frac{\partial G(u,v)}{\partial u}$ exists and is strictly positive on $\mathbb{R}^2$.*

(iii) *For each $t \in \mathbb{R}$ and fixed γ_0 and δ_0, there exists function $q(\cdot, \cdot)$ with $E[q(X_1, X)^2] < \infty$ that dominates $\frac{\partial G(x_1+t,x)}{\partial t}$.*

ASSUMPTION SB.2. *Denote the joint c.d.f. of unobserved payoff components U and V as $G(\cdot, \cdot)$, where $G_v(\cdot)$ is the marginal c.d.f. of V. Let $\mathcal{G}$ be the class of distributions of errors $g(\cdot, \cdot)$ and covariates $f(\cdot, \cdot)$ that satisfy the assumptions of Theorem 2.1 in the main text and the following additional conditions:*

(i) *There exists a nondecreasing function[1] $\nu(\cdot)$ such that for any $|t| < \infty$,*

$$\lim_{c\to+\infty} \frac{1}{\nu(c)} \sup_{f,g\in\mathcal{G}} E_{f,g}\left[\left(\frac{\partial G(X_1+t,X)}{\partial t}\right)^2 G(X_1+t,X)^{-1}\right.$$
$$\left. \times \left(G_v(X) - G(X_1+t,X)\right)^{-1} \middle| |X_1|, |X| < c\right] < \infty.$$

(ii) *There exists a nonincreasing function $\beta(\cdot)$ such that for any $|t| < \infty$,*

$$\lim_{c\to+\infty} \beta(c) \sup_{f,g\in\mathcal{G}} E_{f,g}\left[\log\left(G(X_1+t,X)\right.\right.$$
$$\left.\left. \times \left(G_v(X) - G(X_1+t,X)\right)\right) \middle| |X_1|, |X| > c\right]^{-1} < \infty,$$

---

[1]We use the same c to trim the support of covariates X and X_1 for notational and algebraic convenience only. Our analysis has a straightforward extension to the case where the relative tail behaviors of X_1 and X are different. In that case $\nu(\cdot)$ will be a function of two arguments.

where $E_{f,g}$ denotes the expectation operator with respect to densities f and g. This assumption allows the inverse joint cumulative distribution function to be nonintegrable in the $\mathbb{R}^2$ plane (its improper integral diverges). It is, however, integrable on any square with finite edge and its integral can be expressed as a function of the length of the edge. The following theorem outlines our main result on the optimal rate for the parameter in the triangular model.

THEOREM SB.1. *Suppose that Assumptions* SB.1 *and* SB.2 *hold. Suppose that $c_n \to \infty$ is a sequence such that $\frac{n\beta^2(c_n)}{\nu(c_n)} = O(1)$ with $n/\nu(c_n) \to \infty$. Then, for this sequence, $\sqrt{\frac{n}{\nu(c_n)}}$ is the optimal rate for the estimator for parameter α_0 in the model described in* (2.3) *in the main paper.*

PROOF. We first prove the following lemma.

LEMMA SB.1. *Suppose that the choice probability functions are estimated via an orthogonal sequence*

$$\mathcal{H}^{(K)}(\cdot, \cdot) = \big(\mathcal{H}_k(x_1, x)\big)_{k=0}^{K}$$

and

$$\inf_{\mu \in \mathbb{R}^K} \big\| P^{y_1 y_2}(x_1, x) - \mu' \mathcal{H}^{(K)}(x_1, x) \big\| = O\big(K^{-r}\big)$$

with

$$\widehat{P}_n^{y_1 y_2}(x_1, x) = \widehat{\mu}' \mathcal{H}^{(K)}(x_1, x), \quad y_1, y_2 \in \{0, 1\},$$

where $\widehat{\mu}$ are the estimated coefficients of the OLS regression of $y_1 y_2$ on the elements of $\mathcal{H}^{(K)}(x_1, x)$. We assume that $K \to \infty$ as $n \to \infty$ such that $n/(K \log n) \to \infty$. The estimator is then constructed by defining the likelihood with support restricted to the set $\{|x_1|, |x| \le c_n\}$. Suppose that a sequence c_n is selected such that $\nu(c_n)/n \to 0$, $K^r/\nu(c_n) \to 0$, and $\nu(c_n)K^2/n \to \infty$. Then, for any sequence $\widehat{\alpha}_n$ with the function $\widehat{l}(\alpha)$ corresponding to the maximand of (SB.2) *such that*

$$\widehat{l}_{K,c_n}(\widehat{\alpha}_{0,n}) \ge \sup_\alpha \widehat{l}_{K,c_n}(\alpha) - o_p\bigg(\sqrt{\frac{\nu(c_n)}{n}}\bigg),$$

we have

$$\sqrt{\frac{n}{\nu(c_n)}}\big|\widehat{\alpha}_{0,n}^* - \alpha_0\big| = O_p(1).$$

PROOF. We introduce the "uncensored" objective function

$$l(\alpha; y_1, y_2, x_1, x) = y_1 y_2 \log \widehat{P}_n^{11}(x_1 + \alpha, x) + (1 - y_1)y_2 \log \widehat{P}_n^{01}(x_1 + \alpha, x)$$

with $Q(\alpha) = E[l(\alpha; y_1, y_2, x_1, x)]$ and $\widehat{P}_n^{11}$ as defined on the following page. Denote

$$\widehat{l}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} l(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i).$$

Also use the censoring function $\omega_n(\cdot) = \mathbf{1}\{|\cdot| \leq c_n\}$:

$$\ell(\alpha; y_1, y_2, x_1, x) = y_1 y_2 \omega_n(x_1 + \alpha)\omega_n(x) \log P^{11}(x_1 + \alpha, x)$$
$$+ (1 - y_1)y_2 \omega_n(x_1 + \alpha)\omega_n(x) \log P^{01}(x_1 + \alpha, x)$$

and

$$\hat{\ell}(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \ell(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i).$$

Now consider the decomposition of the objective function

$$\hat{l}(\alpha) - \hat{\ell}(\alpha_0) = R_1 + R_2 + R_3 + R_4 + R_5 + R_6,$$

where

$$R_1 = \hat{l}(\alpha) - \hat{\ell}(\alpha) - E[\hat{l}(\alpha)] + E[\hat{\ell}(\alpha)],$$
$$R_2 = \hat{\ell}(\alpha) - \hat{\ell}(\alpha_0) - E[\hat{\ell}(\alpha)] + E[\hat{\ell}(\alpha_0)],$$
$$R_3 = E[\hat{l}(\alpha)] - E[\hat{\ell}(\alpha)], \qquad R_4 = E[\hat{\ell}(\alpha)] - Q(\alpha)$$
$$R_5 = -E[\hat{\ell}(\alpha_0)] + Q(\alpha_0), \qquad R_6 = Q(\alpha) - Q(\alpha_0).$$

Term R_1. For convenience, we introduce new notation denoting

$$p^{Kk}(z) = \omega_n(x_1)\omega_n(x)\big[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\big]\big[\mathcal{H}_{l_2}(c_n) - \mathcal{H}_{l_2}(x)\big]$$

and introduce vectors $p^K(z) = (p^{K1}(z), \ldots, p^{KK}(z))'$. Also let $d_i^{00} = (1 - y_{1i})(1 - y_{2i})$ and $d^{00} = (d_1^{00}, \ldots, d_n^{00})'$. Let $\Delta(z) = E[d^{00}|z]$ and $\Delta = (\Delta(z_1), \ldots, \Delta(z_n))'$. We can project this function of z on K basis vectors of the sieve space. Let β be the vector of coefficients of this projection. As demonstrated in Newey (1997), for $P = (p^K(z_1), \ldots, p^K(z_n))'$ and $\hat{Q} = P'P/n$,

$$\|\hat{Q} - Q\| = O_p\left(\sqrt{\frac{K}{n}}\right),$$

where his $\zeta_0(K) = C$ and Q is nonsingular by assumption with the smallest eigenvalue bounded from below by some constant $\underline{\lambda} > 0$. Hence the smallest eigenvalue of $\hat{Q}$ will converge to $\underline{\lambda} > 0$. Following Newey (1997), we use the indicator 1_n to indicate the cases where the smallest eigenvalue of $\hat{Q}$ is above $\frac{\lambda}{2}$ to avoid singularities. We also introduce

$$m^{Kk}(z) = \omega_n(x_1)\omega_n(x)\big[\mathcal{H}_{l_1}(x_1) - \mathcal{H}_{l_1}(-c_n)\big]\big[\mathcal{H}_{l_2}(x) - \mathcal{H}_{l_2}(-c_n)\big].$$

We then can write the estimator

$$\hat{P}^{11}(x_1, x) = m^K(z)'\hat{Q}^{-1}Pd^{00}/n.$$

Note that

$$m^{K'}(z)(\hat{\beta} - \beta) = m^{K'}(z)(\hat{Q}^{-1}P'(d^{00} - \Delta)/n + \hat{Q}^{-1}P'(\Delta - P\beta)/n). \tag{SB.1}$$

We can evaluate the component in the second term as

$$\|P(\Delta - P\beta)/n\| = \sqrt{\sum_{k=1}^{K}\left(\frac{1}{n}\sum_{i=1}^{n}p^{Kk}(z_i)\big(\Delta(z_i) - p^K(z_i)'\beta\big)\right)^2}$$

$$\leq \sqrt{KCK^{-2r}} = O(K^{\frac{1}{2}-r})$$

provided our assumption regarding the sieve space (Assumption SA.2 in Supplemental Material A). As we demonstrate, this result allows us to concentrate on the first term and ignore the second one. For the first term in (SB.1), we can use the result that the smallest eigenvalue of $\hat{Q}$ is converging to $\underline{\lambda} > 0$. Then application of the Cauchy–Schwarz inequality leads to

$$\big|m^{K\prime}(z)\hat{Q}^{-1}P'(d^{00} - \Delta)\big| \leq \big\|\hat{Q}^{-1}m^K(z)\big\|\big\|P'(d^{00} - \Delta)\big\|.$$

Then

$$\big\|\hat{Q}^{-1}m^K(z)\big\| \leq \frac{C}{\underline{\lambda}}\sqrt{K}$$

and

$$\|P'(d^{00} - \Delta)\| = \sqrt{\sum_{k=1}^{K}\left(\sum_{i=1}^{n}p^{Kk}(z_i)\big(d_i^{00} - \Delta(z_i)\big)\right)^2}$$

$$\leq \sqrt{K}\max_{k}\left|\sum_{i=1}^{n}p^{Kk}(z_i)\big(d_i^{00} - \Delta(z_i)\big)\right|.$$

Thus,

$$\big|m^{K\prime}(z)\hat{Q}^{-1}P'(d^{00} - \Delta)/n\big| \leq \frac{CK}{\underline{\lambda}}\max_{k}\left|\frac{1}{n}\sum_{i=1}^{n}p^{Kk}(z_i)\big(d_i^{00} - \Delta(z_i)\big)\right|.$$

Denote

$$\mu_n = \mu\frac{n^{\delta/2}}{\sqrt{nK}} = \gamma_n/K$$

for any $\delta \in (0, 1]$. Next we adapt the arguments for proving Theorem 37 in Pollard (1984) to provide the bound for

$$P\left(\sup_{z}\frac{1}{n}\big\|m^{K\prime}(z)\hat{Q}^{-1}P'(d^{00} - \Delta)\big\| > K\mu_n\right).$$

For K nonnegative random variables Y_i, we note that

$$P\left(\max_{i}Y_i > Kc\right) \leq \sum_{i=1}^{K}P(Y_i > c).$$

Using this observation, we find that

$$P\left(\sup_z \frac{1}{n}\left\|m^{K\prime}(z)\hat{Q}^{-1}P'(d^{00}-\varDelta)\right\| > K\mu_n\right) \le \sum_{k=1}^{K} P\left(\left\|\frac{1}{n}\sum_{i=1}^{n} p^{Kk}(z_i)(d_i^{00}-\varDelta(z_i))\right\| > \gamma_n\right),$$

where we used our definition of $\gamma_n = K\mu_n$. This inequality allows us to substitute the tail bound for the class of functions $P_n^{11}(\cdot,\cdot)$ for a tail bound for fixed functions

$$\mathcal{P}_{n,k} = \{p^{Kk}(\cdot)(d^{00}-\varDelta(\cdot))\}.$$

Then we can apply the inequality from Theorem 37 in Pollard (1984) to obtain

$$P\left(\frac{1}{n}\left\|\sum_{i=1}^{n} p^{Kk}(z_i)(d_i^{00}-\varDelta(z_i))\right\| > \gamma_n\right) \le 2\exp\left(-\frac{2n\gamma_n^2}{C^2} + A'\gamma_n^{-\gamma'}\right).$$

As a result, we find that

$$P\left(\sup_z \frac{1}{n}\left\|m^{K\prime}(z)\hat{Q}^{-1}P'(d^{00}-\varDelta)\right\| > K\mu_n\right) \le 2K\exp\left(-\frac{2n\gamma_n^2}{C^2} + A'\gamma_n^{-\gamma'}\right).$$

Then, provided that $n/\log K \to \infty$ and $\gamma' < 1$, we prove that the right-hand side of this inequality converges to 0. This means that

$$\sup_{(x_1,x)\in\mathcal{X}} \left\|\hat{P}^{11}(x_1,x) - \text{proj}\big(P^{11}(x_1,x)|H_K\big)\right\| = o_p\big(n^{\frac{\gamma}{2}-\frac{1}{2}}\big).$$

From the second term, we provide the evaluation

$$\sup_{P^{11}\in H} \sup_{x_1,x} \left\|\text{proj}\big(P^{11}(x_1,x)|H_K\big) - P^{11}(x_1,x)\right\| = O\big(K^{-r}\big).$$

Therefore, if $K^r/n^{(1-\delta)/2} \to \infty$, then the "bias" term will be negligible. Next we note that similar evaluations can be provided for P^{01}. As the density of (U,V) is strictly positive on $\mathbb{R}^2$, the probabilities are bounded away from zero on any bounded subset of $\mathbb{R}^2$, and we can make the same evaluations for $\log P^{11}(\cdot)$ and $\log P^{01}(\cdot)$. As a result, we can attain the rate

$$\sup_\alpha \left|\hat{l}(\alpha) - \hat{\ell}(\alpha) - E\big[\hat{l}(\alpha)\big] + E\big[\hat{\ell}(\alpha)\big]\right| = o_p\big(n^{-(1-\delta)/2}\big).$$

Term R_3. Consider the approximation bias term. Note that we can express

$$E\big[\hat{l}(\alpha)\big] = E\big[\omega_n(x_1+\alpha)\omega_n(x)\big(P^{11}(x_1+\alpha,x)\log\hat{P}_n^{11}(x_1+\alpha,x)$$
$$+ P^{01}(x_1+\alpha,x)\log\hat{P}_n^{01}(x_1+\alpha,x)\big)\big].$$

Similarly, we can express

$$E\big[\hat{\ell}(\alpha)\big] = E\big[\omega_n(x_1+\alpha)\omega_n(x)\big(P^{11}(x_1+\alpha,x)\log P^{11}(x_1+\alpha,x)$$
$$+ P^{01}(x_1+\alpha,x)\log P^{01}(x_1+\alpha,x)\big)\big].$$

One can attain a uniform rate

$$\sup_{x_1, x} \left\| \hat{P}_n^{11}(x_1 + \alpha, x) - P^{11}(x_1 + \alpha, x) \right\| = O_P\left(\sqrt{\frac{K}{n}} + K^{-r} \right),$$

given the quality of approximation by selected sieves. We can then evaluate the entire term

$$|R_3| = O\left(\sqrt{\frac{K}{n}} + K^{1-r} \right).$$

Terms R_4 and R_5. Consider term R_4. We can evaluate this term as

$$\left| E[\hat{\ell}(\alpha)] - Q(\alpha) \right| \leq 4 \left| \int_{-\infty}^{c_n} \int_{-\infty}^{c_n} P^{11}(x_1 + \alpha, x) \log P^{11}(x_1 + \alpha, x) f(x_1, x) \, dx_1 \, dx \right|.$$

We can then apply the Cauchy–Schwarz inequality and continue evaluation as

$$\left| E[\hat{\ell}(\alpha)] - Q(\alpha) \right| \leq 4E[y_1 y_2] \left| \int_{-\infty}^{c_n} \int_{-\infty}^{c_n} \log P^{11}(x_1 + \alpha, x) f(x_1, x) \, dx_1 \right|$$

$$\leq C\beta(c_n)$$

from Assumption SB.2.

Term R_2. We use the following assumption regarding the population likelihood function.

ASSUMPTION SB.3. *The population likelihood function $Q(\cdot)$ is twice continuously differentiable and uniquely maximized at α_0 with a negative definite Hessian.*

Consider the class of functions indexed by $\alpha \in A$ such that given

$$\ell(\alpha, y_1, y_2, x_1, x) = \left[y_1 y_2 \log P^{11}(x_1 + \alpha, x) + (1 - y_1) y_2 \log P^{01}(x_1 + \alpha, x) \right]$$

$$\times \omega_n(x_1 + \alpha) \omega_n(x),$$

then

$$\mathcal{F}_{n,\delta} = \left\{ f = \ell(\alpha, \cdot) - \ell(\alpha_0, \cdot), |\alpha - \alpha_0| \leq \delta \right\}.$$

Provided that the density of errors is twice differentiable in mean square with bounded mean-square derivatives, there exist bounded functions $\dot{P}^{11}$ and $\dot{P}^{01}$ such that functions in class $\mathcal{F}_{n,\delta}$ have envelope

$$F_{n,\delta} = \mathbf{1}\left\{ |x_1 + \alpha_0| \leq c_n + \delta \right\} \omega_n(x)$$

$$\times \left[\frac{y_1 y_2 \dot{P}^{11}}{P^{11}} + \frac{(1 - y_1) y_2 \dot{P}^{01}}{P^{01}} \right] \delta.$$

Then, by Assumption SB.2, we can evaluate

$$\left(E[F_{n,\delta}^2] \right)^{1/2} = O\left(\nu(c_n)^{1/2} \delta \right).$$

Consider the reparametrization of the model $\alpha = \alpha_0 + \frac{h}{r_n}$ for a sequence $r_n \to \infty$. Take $h \in [0, \eta r_n]$ for some large η and split the interval $[0, \eta r_n]$ into "shells" $S_{n,j} = \{h : 2^{j-1} < |h| < 2^j\}$. Suppose that $\hat{h}$ is the maximizer for $\hat{l}(\alpha_0 + \frac{h}{r_n})$. Then if $|\hat{h}| > 2^M$ for some M, then $\hat{h}$ belongs to $S_{n,j}$ with $j \geq M$. As a result

$$P(|\hat{h}| > 2^M) \leq \sum_{j \geq M, 2^j < \eta r_n} P\left(\sup_{h \in S_{n,j}} \left(\hat{l}\left(\alpha_0 + \frac{h}{r_n}\right) - \hat{l}(\alpha_0) \right) \geq 0 \right).$$

We now use the results from the evaluation of the terms R_1, R_3, R_4, and R_5, taking into consideration that

$$Q(\alpha) - Q(\alpha_0) \leq -H|\alpha - \alpha_0|^2$$

for some $H > 0$ due to the differentiability of $Q(\cdot)$ and the restriction on its Hessian at α_0 in Assumption SB.3. We can evaluate

$$P\left(\sup_{h \in S_{n,j}} \left(\hat{l}\left(\alpha_0 + \frac{h}{r_n}\right) - \hat{l}(\alpha_0) \right) \geq 0 \right)$$

$$\leq P\left(\sup_{h \in S_{n,j}} |R_2| \geq |R_1| + |R_3| + |R_4| + |R_5| + |R_6| \right)$$

$$= P\left(\sup_{h \in S_{n,j}} |R_2| \geq \frac{2^{2j-2}}{r_n^2} + O\left(\sqrt{\frac{K}{n}} + K^{1-r} + \beta(c_n)^{-1} \right) \right),$$

where we use that the difference of absolute values is smaller than the absolute value of the difference. Then we use the Markov inequality to obtain that

$$P\left(\sup_{h \in S_{n,j}} \left(\hat{l}\left(\alpha_0 + \frac{h}{r_n}\right) - \hat{l}(\alpha_0) \right) \geq 0 \right)$$

$$\leq \frac{E\left[\sup_{h \in S_{n,j}} \left| \hat{\ell}\left(\alpha_0 + \frac{h}{r_n}\right) - \hat{\ell}(\alpha_0) - E\left[\hat{\ell}\left(\alpha_0 + \frac{h}{r_n}\right) \right] + E[\hat{\ell}(\alpha_0)] \right| \right]}{\frac{2^{2j-2}}{r_n^2} + O\left(\sqrt{\frac{K}{n}} + K^{1-r} + \beta(c_n)^{-1} \right)}.$$

Using empirical process notation, we define the covering integral as

$$J(\delta, F) = \sup_Q \int_0^\delta \sqrt{1 + \log N\left(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q) \right)} \, d\varepsilon,$$

where Q is the probability measure, $\mathcal{F}$ is a class of functions with the envelope F, and $N(\cdot)$ is the covering number of the considered class. Provided the finiteness of the cov-

ering integral of the class $\mathcal{F}_{n,\delta}$, we can use the maximum inequality to evaluate

$$E\left[\sup_{h\in S_{n,j}}\sqrt{n}\left|\hat{\ell}\left(\alpha_0+\frac{h}{r_n}\right)-\hat{\ell}(\alpha_0)-E\left[\hat{\ell}\left(\alpha_0+\frac{h}{r_n}\right)\right]+E[\hat{\ell}(\alpha_0)]\right|\right]$$

$$\le J(1,F_{n,h/r_n})E[F^2_{n,h/r_n}]^{1/2}=O\left(\nu(c_n)^{1/2}\frac{2^j}{r_n}\right).$$

Assuming that $r_n\beta(c_n)^{-1}=o(1)$, $r_n\sqrt{K/n}=o(1)$, and $r_nK^{-(d+1)/2}\to 0$, then

$$P\left(\sup_{h\in S_{n,j}}\left(\hat{l}\left(\alpha_0+\frac{h}{r_n}\right)-\hat{l}(\alpha_0)\right)\ge 0\right)\le O\left(2^{-j+2}r_n\sqrt{\frac{\nu(c_n)}{n}}\right).$$

This implies that

$$P(|\hat{h}|>2^M)\le O\left(2^{-M+3}r_n\sqrt{\frac{\nu(c_n)}{n}}\right).$$

The right-hand side converges to 0 for $M\to\infty$ if $r_n=\sqrt{\frac{n}{\nu(c_n)}}$. $\square$

SB.1.1 *Proof of convergence rates in triangular model*

First, consider the evaluation from the proof of Lemma SB.1:

$$P\left(\sup_{h\in S_{n,j}}\left(\hat{l}\left(\alpha_0+\frac{h}{r_n}\right)-\hat{l}(\alpha_0)\right)\ge 0\right)$$

$$\le\frac{E\left[\sup_{h\in S_{n,j}}\left|\hat{\ell}\left(\alpha_0+\frac{h}{r_n}\right)-\hat{\ell}(\alpha_0)-E\left[\hat{\ell}\left(\alpha_0+\frac{h}{r_n}\right)\right]+E[\hat{\ell}(\alpha_0)]\right|\right]}{\frac{2^{2j-2}}{r_n^2}+O\left(\sqrt{\frac{K}{n}}+K^{-(d+1)/2}+\beta(c_n)^{-1}\right)}.$$

Using the maximum inequality as before, we can conclude that the ratio can be evaluated as

$$P\left(\sup_{h\in S_{n,j}}\left(\hat{l}\left(\alpha_0+\frac{h}{r_n}\right)-\hat{l}(\alpha_0)\right)\ge 0\right)\le O\left(2^{-j+1}\frac{\nu(c_n)^{1/2}r_n}{\sqrt{n}}\right).$$

We note that evaluation here is different, because, unlike in Lemma SB.1, here we allow $r_n\beta(c_n)=O(1)$. This allows us to obtain

$$P\left(\sqrt{\frac{n}{\nu(c_n)}}|\hat{h}|>2^M\right)\le O\left(2^{-M+2}r_n\sqrt{\frac{\nu(c_n)}{n}}\right).$$

Thus, if $L=2^M$, then

$$P\left(\sqrt{\frac{n}{\nu(c_n)}}|\hat{h}|>L\right)\le O\left(\frac{4}{L}r_n\sqrt{\frac{\nu(c_n)}{n}}\right).$$

Provided that we choose $r_n\sqrt{\frac{\nu(c_n)}{n}} = 1$, we assure that for the maximal risk,

$$\lim_{L\to\infty} \limsup_{n\to\infty} R\left(\alpha_0 + \frac{\hat{h}}{r_n}, r_n, L\right) = 0.$$

This means that r_n is the upper rate.

To derive the lower convergence rate, we use the result from Koroselev and Tsybakov (1993). Denote the likelihood ratio as $\Lambda(P_1, P_2) = \frac{dP_{P_1}^n}{dP_{P_2}^n}$. Then the following lemma is the result given in Koroselev and Tsybakov (1993).

LEMMA SB.2. *Suppose that $\alpha_0^1 = \alpha(P_1)$ and $\alpha_0^2 = \alpha(P_2)$, and let $\lambda > 0$ be such that*

$$P_{P_2}\big(\Lambda(P_1, P_2) > \exp(-\lambda)\big) \geq p > 0$$

and $|\alpha_0^1 - \alpha_0^2| \geq 2s_n$. Then, for any estimator $\hat{\alpha}_{0,n}$, we have

$$\max_{P_1,P_2} P\big(|\hat{\alpha}_{0,n} - \alpha_0| > s_n\big) \geq p\exp(-\lambda/2).$$

We can now use this lemma to derive the following result regarding the lower rate for the estimator of interest.

The log-likelihood function of the model is

$$n\hat{\mathcal{L}}(\alpha) = n\hat{\ell}(\alpha) + n\hat{e}(\alpha)$$

with

$$\hat{e}(\alpha) = \frac{1}{n}\sum_{i=1}^n \big\{y_{1i}\log P^{11}(x_{1i} + \alpha, x_i) + (1 - y_{1i})\log P^{01}(x_{1i} + \alpha, x_i)\big\}$$

$$\times y_{2i}\mathbf{1}\big\{|x_{1i}| > c_n, |x_i| > c_n\big\}.$$

Note that we use the same distribution of covariates x_1 and x. For $c_n \to \infty$, pick[2]

$$P_2(\cdot, \cdot) = P(\cdot, \cdot) \quad\text{and}\quad P_1(\cdot, \cdot) = P(\cdot, \cdot)\omega_n(\cdot)\omega_n(\cdot).$$

Following from our previous analysis for such choices of $P_1(\cdot)$ and $P_2(\cdot)$, the corresponding likelihood maximizers satisfy

$$|\alpha_1 - \alpha_2| = O\big(\beta(c_n)\big).$$

We can then express

$$\Lambda(P_1, P_2) = \exp\big(n\hat{\mathcal{L}}_1(\alpha_1) - n\hat{\mathcal{L}}_2(\alpha_2)\big)$$

$$= \exp\big(n\hat{\ell}(\alpha_1) - n\hat{\ell}(\alpha_2) - n\hat{e}(\alpha_2)\big)$$

$$= \exp\big(n\big[\hat{\ell}(\alpha_1) - \hat{\ell}(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2)\big] - n\hat{e}(\alpha_2) - n\big(\ell(\alpha_2) - \ell(\alpha_1)\big)\big).$$

---

[2]The selected P_1 may not be a probability measure; however, it bears the properties of the measure, such that characteristics of the measure such as a Radon–Nykodim derivative are still well defined.

We note that

$$\hat{\ell}(\alpha_1) - \hat{\ell}(\alpha_2) - \ell(\alpha_1) + \ell(\alpha_2) = o_p(1)$$

and

$$\hat{e}(\alpha_2) = o_p(1).$$

As a result, the last term dominate as $n \to \infty$. Then $\log \Lambda(P_1, P_2)$ is bounded from below as n approaches infinity if and only if $n(\ell(\alpha_2) - \ell(\alpha_1))$ is bounded. We note that α_1 maximizes $\ell(\alpha)$. This means that

$$\ell(\alpha_2) - \ell(\alpha_1) = -\frac{1}{2}H(c_n)(\alpha_2 - \alpha_1)^2 + o(|\alpha_2 - \alpha_1|).$$

Invoking the Cauchy–Schwarz inequality, we can evaluate $H(c_n) = O(\nu(c_n)^{-1})$. As a result, we find that

$$n[\ell(\alpha_2) - \ell(\alpha_1)] = O\left(\frac{n\beta(c_n)^2}{\nu(c_n)}\right).$$

This means that $\frac{n\beta(c_n)^2}{\nu(c_n)} = O(1)$, suggesting that for large n, there exists a lower bound on the likelihood ratio. By invoking Lemma SB.2, we obtain the desired result.   □

SB.2. Optimal convergence rate for strategic interaction parameter in a static game model

In this section we show that the proposed estimators achieve the optimal rate of convergence in the triangular model.

Step 1. Consider the family of normalized Hermite polynomials and denote

$$h_l(x) = (\sqrt{2\pi}l!)^{-1/2}e^{-\frac{x^2}{4}}H_l(x),$$

where $H_l(\cdot)$ is the lth degree Hermite polynomial. Also denote

$$\mathcal{H}_l(x) = \int_{-\infty}^{x} h_l(z)\,dz.$$

We note that this sequence is orthonormal for the inner product defined as

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x)g(x)\,dx.$$

We take the sequence $c_n \to \infty$, define the function $\omega_n(x) = \mathbf{1}\{|x| \le c_n\}$, and estimate the probability that both indicators are equal to 0 ($y_1 = y_2 = 0$) as

$$\hat{\mathcal{P}}_n^{00}(x_1, x) = \sum_{l_1,l_2=1}^{K(n)} a_{l_1 l_2}\omega_n(x_1)\big[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\big]\omega_n(x)\big[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x)\big].$$

The estimates can be obtained via a regression of

$$\omega_n(x_1)\big[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1)\big]\omega_n(x)\big[\mathcal{H}_{l_2}(c_n) - \mathcal{H}_{l_2}(x)\big]$$

on the indicators $(1 - y_1)(1 - y_2)$. Then the estimator for the joint density of errors can be obtained from the regression coefficients as

$$\hat{g}_n(x_1, x) = \sum_{l_1, l_2 = 1}^{K(n)} \hat{a}_{l_1 l_2}\omega_n(x_1)h_{l_1}(x_1)\omega_n(x)h_{l_2}(x).$$

Step 2. Using the estimator for the density, we compute the fitted values for conditional probabilities of $y_1 = y_2 = 1$ and $y_1 = 0$, $y_2 = 1$ as

$$\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x) = \sum_{l_1, l_2 = 1}^{K(n)} \hat{a}_{l_1 l_2}\omega_n(x_1 + \alpha)\big[\mathcal{H}_{l_1}(x_1 + \alpha) - \mathcal{H}_{l_1}(-c_n)\big]\omega_n(x)\big[\mathcal{H}_{l_1}(x) - \mathcal{H}_{l_1}(-c_n)\big]$$

and

$$\hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x) = \sum_{l_1, l_2 = 1}^{K(n)} \hat{a}_{l_1 l_2}\omega_n(x_1 + \alpha)\big[\mathcal{H}_{l_1}(c_n) - \mathcal{H}_{l_1}(x_1 + \alpha)\big]\omega_n(x)\big[\mathcal{H}_{l_1}(x) - \mathcal{H}_{l_1}(-c_n)\big].$$

Using these fitted probabilities, we can form the conditional log-likelihood function

$$l(\alpha; y_1, y_2, x_1, x) = y_1 y_2 \omega_n(x_1 + \alpha)\omega_n(x)\log \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)$$
$$+ (1 - y_1)y_2\omega_n(x_1 + \alpha)\omega_n(x)\log \hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x).$$

Then we can express the empirical score as

$$s(\alpha; y_1, y_2, x_1, x) = \left[\frac{y_1 y_2}{\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)} - \frac{(1 - y_1)y_2}{\hat{\mathcal{P}}_n^{01}(x_1 + \alpha, x)}\right]\frac{\partial \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\partial \alpha}\omega_n(x_1 + \alpha)\omega_n(x).$$

This expression can be rewritten as

$$s(\alpha; y_1, y_2, x_1, x) = \frac{\omega_n(x_1 + \alpha)\omega_n(x)y_2}{\hat{\mathcal{P}}_n^{11}(c_n, x)} \frac{y_1 - \dfrac{\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\hat{\mathcal{P}}_n^{11}(c_n, x)}}{\left(1 - \dfrac{\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\hat{\mathcal{P}}_n^{11}(c_n, x)}\right)\dfrac{\hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\hat{\mathcal{P}}_n^{11}(c_n, x)}}\frac{\partial \hat{\mathcal{P}}_n^{11}(x_1 + \alpha, x)}{\partial \alpha}.$$

Setting the empirical score equal to zero, we obtain the estimator for α_0 as

$$\hat{\alpha}_n^* = \operatorname*{argmax}_{\alpha} \frac{1}{n}\sum_{i=1}^{n} l(\alpha; y_{1i}, y_{2i}, x_{1i}, x_i). \tag{SB.2}$$

SB.3. EXAMPLES OF CONVERGENCE RATES FOR COMMON CLASSES OF DISTRIBUTIONS

Here we illustrate how rates of convergence can depend on relative tail conditions by considering particular parametric distributions of observed and unobserved variables.

To evaluate function $\nu(\cdot)$, we consider the one-dimensional case. Let $F(\cdot)$ be the c.d.f. of the errors and let $\phi(\cdot)$ be the probability density function (p.d.f.) of the covariates. We note that in the one-dimensional case, $\frac{\partial}{\partial t} G(x_1 + t, x)$ corresponds to $F(x_1 + t)$ and $G(x_1 + t, x)(G_v(x) - G(x_1 + t, x))$ corresponds to $F(x_1 + t)(1 - F(x_1 + t))$. Thus, given that we evaluate

$$\nu(c) = O\left(E\left[\left(\frac{\partial}{\partial t} G(X_1 + t, X) \right)^2 G(X_1 + t, X)^{-1} \left(G_v(X) - G(X_1 + t, X) \right)^{-1} \Big| |X_1|, |X| < c \right] \right).$$

In the one-dimensional case this leads to

$$\nu(c) = O\left(\int_{-c}^{c} \frac{F(x_1 + t)^2}{F(x_1 + t)\left(1 - F(x_1 + t)\right)} \phi(x_1)\, dx_1 \right).$$

For our considered distributions, $F(x_1 + t)$ is symmetric in $x_1 + t$, meaning that $\nu(c)$ will be majorized by the choice $t = 0$. Next we notice that

$$\frac{F(x_1)^2}{F(x_1)\left(1 - F(x_1)\right)} = \frac{F(x_1)}{1 - F(x_1)} \leq \frac{1}{1 - F(x_1)}.$$

Note that for our considered distributions, the right and left tails behave equivalently and, thus, we can consider integration only over positive x_1. We conclude that we can evaluate

$$\nu(c) = O\left(\int_0^c \frac{\phi(x_1)}{1 - F(x_1 + t)}\, dx_1 \right).$$

We evaluate the term of interest as

$$\int_0^c \frac{\phi(x)}{1 - F(x)}\, dx = \int_0^c \frac{e^x}{1 + e^x}\, dx.$$

A change of variables $z = e^x$ allows us to rewrite this expression as

$$\int_1^{e^c} \frac{dz}{1 + z} = O(c).$$

Given that we have a two-dimensional distribution, we can select $\nu(c) = c^2$. Next we evaluate function $\beta(\cdot)$, whose leading term can be represented as

$$\int_c^{\infty} \log\left((1 + e^x)^{-1}\right) \frac{e^x}{\left(1 + e^x\right)^2}\, dx = O(e^{-c}).$$

Therefore, we can select $\beta(c) = e^{-c}$ and the optimal rate will be $\sqrt{n/c_n^2}$ with $c_n e^{c_n}/n = O(1)$. For example, we can select $c_n = \delta\sqrt{\log n}$ for some $0 < \delta < 1$, delivering convergence rate $\sqrt{n/\log n}$. Similar arguments can be used for other distributions. Examples are in Table SB.1.

TABLE SB.1. Elements of the optimal rate for various distributions.

	Covariate Distribution	
Error Distribution	Logistic	Normal
Logistic	$\nu(c) = c^2,\ \beta(c) = e^{-c}$ Rate $\sim \sqrt{n/\log n}$	$\nu(c) = 1,\ \beta(c) = e^{-c^2}$ Rate $\sim \sqrt{n}$
Normal	$\nu(c) = e^{c^2},\ \beta(c) = c^2 e^{-c}$ Rate $\sim n^{1/4}$	$\nu(c) = c^4,\ \beta(c) = c e^{-c^2}$ Rate $\sim \sqrt{n}/\log n$

SB.4. SEMIPARAMETRIC EFFICIENCY BOUND IN THE TRIANGULAR MODEL WITH INCOMPLETE INFORMATION

The semiparametric efficiency bound provides the minimum variance for the finite-dimensional parameters over admissible sets of nonparametric components of the model. We take the triangular model with incomplete information constructed in Khan and Nekipelov (2010). We follow Ai and Chen (2003); note that the model is represented by a system of semiparametric conditional moment equations:

$$P_{11}(x_1, x) = E[y_1 y_2 | x_1, x] = \int \mathbf{1}\left\{x_1 - u + \alpha \Phi\left(\frac{x-v}{\sigma}\right) > 0\right\} \Phi\left(\frac{x-v}{\sigma}\right) g(u, v)\, du\, dv,$$

$$P(x_1, x) = E[y_2 | x_1, x] = \int \Phi\left(\frac{x-v}{\sigma}\right) g_v(v)\, dv, \tag{SB.3}$$

$$Q(x_1, x) = E[y_1 | x_1, x] = \int \mathbf{1}\left\{x_1 - u + \alpha \Phi\left(\frac{x-v}{\sigma}\right) > 0\right\} g(u, v)\, du\, dv.$$

These equations fully characterize the conditional distribution of the outcome variables. We can rewrite this system of equations in an equivalent form as

$$m_1(x_1, x; \alpha, g) = E\left[y_1 y_2 - \int \mathbf{1}\left\{x_1 - u + \alpha \Phi\left(\frac{x-v}{\sigma}\right) > 0\right\}\right.$$
$$\left. \times \Phi\left(\frac{x-v}{\sigma}\right) g(u, v)\, du\, dv \Big| x_1, x\right] = E\left[\rho_1(y, x; \alpha, g)\Big| x_1, x\right],$$
$$m_2(x_1, x; \alpha, g) = E\left[y_1 - \int \mathbf{1}\left\{x_1 - u + \alpha \Phi\left(\frac{x-v}{\sigma}\right) > 0\right\} g(u, v)\, du\, dv\right] \tag{SB.4}$$
$$= E\left[\rho_2(y, x; \alpha, g)\Big| x_1, x\right],$$
$$m_3(x_1, x; \alpha, g) = E\left[y_2 - \int \Phi\left(\frac{x-v}{\sigma}\right) g_v(v)\, dv \Big| x_1, x\right] = E\left[\rho_3(y, x; \alpha, g) | x_1, x\right].$$

Consider the derivatives of these moment equations with respect to parameter α:

$$\frac{dm_1}{d\alpha} = -\int \Phi\left(\frac{x-v}{\sigma}\right)^2 \frac{\partial}{\partial v} G\left(x_1 + \alpha\Phi\left(\frac{x-v}{\sigma}\right), v\right) du\, dv,$$

$$\frac{dm_2}{d\alpha} = -\int \Phi\left(\frac{x-v}{\sigma}\right) \frac{\partial}{\partial v} G\left(x_1 + \alpha\Phi\left(\frac{x-v}{\sigma}\right), v\right) du\, dv,$$

$$\frac{dm_3}{d\alpha} = 0.$$

Then considering the space of densities that are *uniformly manageable,* we take a direction in this space h and

$$\frac{dm_1}{dg}[h] = -\int \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\}\Phi\left(\frac{x-v}{\sigma}\right) h(u,v)\, du\, dv,$$

$$\frac{dm_2}{dg}[h] = -\int \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\} h(u,v)\, du\, dv,$$

$$\frac{dm_3}{dg}[h] = -\int \Phi\left(\frac{x-v}{\sigma}\right) h_v(v)\, dv.$$

We introduce the vector with elements

$$\psi_1(x_1, x, u, v) = \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\}\left(\Phi\left(\frac{x-v}{\sigma}\right) g(u,v) - h(u,v)\right),$$

$$\psi_2(x_1, x, u, v) = \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\}(g(u,v) - h(u,v)),$$

$$\psi_3(x_1, x, u, v) = -h(u,v),$$

and denote

$$\zeta_1(x_1, x, u, v) = \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\},$$

$$\zeta_2(x_1, x, u, v) = \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\},$$

$$\zeta_3(x_1, x, u, v) = 1,$$

and

$$\xi_1(x_1, x, u, v) = \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\}\Phi\left(\frac{x-v}{\sigma}\right),$$

$$\xi_2(x_1, x, u, v) = \mathbf{1}\left\{x_1 - u + \alpha\Phi\left(\frac{x-v}{\sigma}\right) > 0\right\},$$

$$\xi_3(x_1, x, u, v) = 1.$$

We express

$$D_h(x_1, x) = \frac{dm}{d\alpha} - \frac{dm}{dg}[h] = \int \Phi\left(\frac{x-v}{\sigma}\right) \psi(x_1, x, u, v)\, du\, dv,$$

which is a linear functional of $h(\cdot, \cdot)$; in fact,

$$D_h(x_1, x) = \int \Phi\left(\frac{x - v}{\sigma}\right) \xi(x_1, x, u, v) g(u, v)\, du\, dv$$

$$- \int \Phi\left(\frac{x - v}{\sigma}\right) \zeta(x_1, x, u, v) h(u, v)\, du\, dv.$$

Next we find the conditional covariance matrix

$$\Sigma(x_1, x) = P_{11}\left(I - \begin{pmatrix} P_{11} & Q & P \\ Q & 1 - \dfrac{Q(1 - Q)}{P_{11}} & \dfrac{PQ}{P_{11}} \\ P & \dfrac{PQ}{P_{11}} & 1 - \dfrac{P(1 - P)}{P_{11}} \end{pmatrix}\right). \tag{SB.5}$$

The semiparametric efficiency bound will be associated with the "least favorable" direction h. To find this direction one needs to solve the minimization problem

$$\min_{h \in \mathcal{G} - g_0} E\left[D_h(X_1, X)' \Sigma^{-1}(X_1, X) D_h(X_1, X)\right].$$

It is convenient to define the least favorable direction as $h = q^2$ to ensure that the solution is positive and also require that $\int q^2(u, v)\, du\, dv = 1$. Then the minimization problem becomes a constrained optimization problem. The considered minimized functional is quadratic and we can express the necessary condition for its minimum as

$$E\left[\Phi\left(\frac{X - v}{\sigma}\right) \zeta(X_1, X, u, v)' \Sigma^{-1}(X_1, X) D_{h^*}(X_1, X)\right] + \lambda = 0,$$

where λ is the Lagrange multiplier and $h^* = q^{*2}$ corresponds to the optimal solution. Finally, we can transform this equation by isolating the terms for h^* and g and introducing notations

$$K(u, v, u', v') = E\left[\Phi\left(\frac{X - v}{\sigma}\right) \Phi\left(\frac{X - v'}{\sigma}\right) \zeta(X_1, X, u, v)' \Sigma^{-1}(X_1, X) \zeta(X_1, X, u', v')\right]$$

and

$$R(u, v, u', v') = E\left[\Phi\left(\frac{X - v}{\sigma}\right) \Phi\left(\frac{X - v'}{\sigma}\right) \zeta(X_1, X, u, v)' \Sigma^{-1}(X_1, X) \xi(X_1, X, u', v')\right].$$

Thus,

$$\int K(u, v, u', v') h^*(u', v')\, du'\, dv' = \lambda + \int R(u, v, u', v') g(u', v')\, du'\, dv'.$$

Given that $K(u, v, u', v')$ is a nonseparable symmetric kernel. Thus it has an infinitely countable set of eigenfunctions with real eigenvalues. The Fredholm integral equation above has a solution. This solution to this equation that is strictly positive and normalizes to 1 yields the semiparametric efficiency bound

$$\Omega = E\left[D_{h^*}(X_1, X)' \Sigma^{-1}(X_1, X) D_{h^*}(X_1, X)\right].$$

<div align="center">

SB.5. Semiparametric efficiency bound in the static game
with incomplete information

</div>

The following result states that the optimal convergence rate for the estimator for the strategic interaction parameters in the incomplete information game is parametric with a limiting normal distribution and the minimum variance of the estimator converging at the parametric rate corresponds to the semiparametric efficiency bound.

The conditional distribution of observed actions is fully characterized by three expectations:

$$E[Y_1|x_1, x_2],$$

$$E[Y_2|x_1, x_2],$$

and

$$E[Y_1 Y_2|x_1, x_2].$$

These expectations characterize the conditional moments that identify the strategic interaction parameters:

$$P_{11}(x_1, x_2) = E[Y_1 Y_2|x_1, x_2] = \int P_1(x_1 - u, x_2 - v) P_2(x_1 - u, x_2 - v) g(u, v)\, du\, dv,$$

$$Q(x_1, x_2) = E[Y_1|x_1, x_2] = \int P_1(x_1 - u, x_2 - v) g(u, v)\, du\, dv,$$

$$P(x_1, x_2) = E[Y_2|x_1, x_2] = \int P_2(x_1 - u, x_2 - v) g(u, v)\, du\, dv.$$

We can rewrite this system of equations in an equivalent form as

$$m_1(x_1, x_2; \alpha, g) = E\left[Y_1 Y_2 - \int P_1(X_1 - u, X_2 - v) P_2(X_1 - u, X_2 - v) \right.$$

$$\left. \times g(u, v)\, du\, dv|x_1, x_2 \right]$$

$$= E\left[\rho_1(Y, X; \alpha, g)|x_1, x \right],$$

$$m_2(x_1, x_2; \alpha, g) = E\left[Y_1 - \int P_1(X_1 - u, X_2 - v) g(u, v)\, du\, dv \right] \qquad \text{(SB.6)}$$

$$= E\left[\rho_2(Y, X; \alpha, g)|x_1, x_2 \right],$$

$$m_3(x_1, x_2; \alpha, g) = E\left[Y_2 - \int P_2(X_1 - u, X_2 - v) g(u, v)\, du\, dv|x_1, x_2 \right]$$

$$= E\left[\rho_3(Y, X; \alpha, g)|x_1, x \right].$$

Under our assumption regarding the distribution of errors η_1 and η_2, equilibrium beliefs are monotone functions of the parameters. Previously, we derived the Jacobi matrix that corresponds to the derivatives of the equilibrium beliefs with respect to the parameters

as

$$J^\alpha = \begin{pmatrix} \dfrac{\partial P_1}{\partial \alpha_1} & \dfrac{\partial P_1}{\partial \alpha_2} \\[2mm] \dfrac{\partial P_2}{\partial \alpha_1} & \dfrac{\partial P_2}{\partial \alpha_2} \end{pmatrix} = \frac{a_1 a_2}{1 + \alpha_1 \alpha_2 a_1 a_2} \begin{pmatrix} P_1/a_2 & \alpha_1 P_1 \\ \alpha_2 P_2 & P_2/a_1 \end{pmatrix},$$

where $a_i = \sigma^{-1} \phi(\Phi^{-1}(P_i))$.

We can express the Jacobi matrix of the moment vector $m(\cdot)$ with respect to the finite-dimensional parameters α_1 and α_2 as

$$\frac{dm(x_1, x_2; \alpha, g)}{d\alpha'} = \int M(x_1 - u, x_2 - v) J^\alpha (x_1 - u, x_2 - v) g(u, v) \, du \, dv,$$

where

$$M = \begin{pmatrix} P_2 & P_1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (\mu_1, \mu_2).$$

Then considering the space of densities that satisfy, we take a direction in this space h and obtain

$$\frac{dm(x_1, x_2; \alpha, g)}{dg}[h] = \int \psi(x_1 - u, x_2 - v) h(u, v) \, du \, dv,$$

where

$$\psi(q_1, q_2) = \big(P_1(q_1, q_2) P_1(q_1, q_2), P_1(q_1, q_2), P_2(q_1, q_2)\big)'.$$

The semiparametric efficiency bound will be associated with a vector of two least favorable directions h_1^* and h_2^* such that h_i^* minimizes

$$E\big[D_{h_i}(X_1, X_2) \Sigma(X_1, X_2)^{-1} D_{h_i}(X_1, X_2)\big],$$

where

$$D_{h_i}(x_1, x_2) = \frac{dm(x_1, x_2; \alpha, g)}{d\alpha_i} - \frac{dm(x_1, x_2; \alpha, g)}{dg}[h_i]$$

and $\Sigma(\cdot, \cdot)$ is determined by (SB.5). We note that $D_{h_i}(x_1, x_2)$ is linear in h_i. We can minimize the considered objective function under the constraint that the solution has to be a density function. This optimization leads us to the expression

$$E\big[\psi(X_1 - u, X_2 - v)' \Sigma(X_1, X_2)^{-1} D_{h_i}(X_1, X_2)\big] + \lambda = 0,$$

where λ is the Lagrange multiplier. We introduce notation

$$K(u, v, u', v') = E\big[\psi(X_1 - u, X_2 - v)' \Sigma(X_1, X_2)^{-1} \psi(X_1 - u', X_2 - v')\big]$$

and

$$R_i(u, v, u', v') = E\big[\psi(X_1 - u, X_2 - v)' \Sigma(X_1, X_2)^{-1} \mu_i(X_1 - u', X_2 - v') J^\alpha(X_1 - u', X_2 - v')\big].$$

Then we can find the least favorable direction for $i = 1, 2$ as a solution to

$$\int K(u, v, u', v') h_i^*(u', v') \, du' \, dv' = \lambda + \int R(u, v, u', v') g(u', v') \, du' \, dv'.$$

The kernel function $K(u, v, u', v')$ is positive, symmetric, nonseparable, and square-integrable. Thus, the Hilbert space $\mathcal{G}$ has an orthonormal basis consisting of the eigenvectors of the integral operator with the kernel $K(u, v, u', v')$, and the solution for h_i^* will be in this basis. The semiparametric efficiency bound isl then be constructed from

$$D_{h^*}(x_1, x_2) = \left(D_{h_1^*}(x_1, x_2), D_{h_2^*}(x_1, x_2) \right)'.$$

We can express the bound as

$$\Omega = E\left[D_{h^*}(X_1, X_2) \Sigma(X_1, X_2)^{-1} D_{h^*}(X_1, X_2) \right]^{-1}.$$

This result is not that surprising in light of the finding in Khan and Nekipelov (2010) that, given that the information for the strategic interaction parameters is positive, the semiparametric efficiency bound will be finite. The efficiency bound for a static two-player game of incomplete information has been analyzed in Aradillas-Lopez (2010) without allowing for player-specific unobserved heterogeneity that is commonly observed by the players. Grieco (2010) allows for individual-specific heterogeneity, but assumes a specific parametric form for both the payoff noise distribution and the distribution of unobserved heterogeneity. We provide the result that parametric inference remains feasible even when the distribution of unobserved heterogeneity remains fully nonparametric. Our efficiency result provides a semiparametric efficiency bound for the generalized class of static games of incomplete information in Bajari, Hong, Krainer, and Nekipelov (2010) as well as in Haile, Hortaçsu, and Kosenok (2008) for the games with quantal response equilibria considered in Palfrey (1985).

References

Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71 (6), 1795–1843. [14]

Aradillas-Lopez, A. (2010), "Semiparametric estimation of a simultaneous game with incomplete information." *Journal of Econometrics*, 157, 409–431. [19]

Bajari, P., H. Hong, M. Krainer, and D. Nekipelov (2010), "Estimating static models of strategic interactions." *Journal of Business & Economic Statistics*, 28, 469–482. [19]

Grieco, P. (2010), "Discrete games with flexible information structures: An application to local grocery markets." Working paper. [19]

Haile, P. A., A. Hortaçsu, and G. Kosenok (2008), "On the empirical content of quantal response equilibrium." *American Economic Review*, 98 (1), 180–200. [19]

Ibragimov, I. and R. Has'minskii (1978), "On the capacity in communication by smooth signals." *Soviet Mathematics. Doklady*, 19, 1043–1047. [1]

Khan, S. and D. Nekipelov (2010), "Semiparametric efficiency in irregularly identified models." Working paper, Duke University. [1, 14, 19]

Koroselev, A. and A. Tsybakov (1993), *Minimax Theory of Image Reconstruction*. Springer, New York. [10]

Newey, W. (1997), "Convergence rates and asymptotic normality for series estimators." *Journal of Econometrics*, 79, 147–168. [4]

Palfrey, T. (1985), "Uncertainty resolution, private information aggregation and the Cournot competitive limit." *Review of Economic Studies*, 52, 69–83. [19]

Pollard, D. (1984), *Convergence of Stochastic Processes*. Springer, New York. [5]