SUPPLEMENT TO "ESTIMATION AND CONFIDENCE REGIONS FOR PARAMETER SETS IN ECONOMETRIC MODELS"

(*Econometrica*, Vol. 75, No. 5, September 2007, 1243–1284)

BY V. CHERNOZHUKOV, E. TAMER, AND H. HONG

In the main text the true probability measure, P, is the nuisance parameter. In this supplementary material we examine which contiguous perturbations of the original fixed P preserve or do not preserve the estimation and coverage properties of the regions constructed in the main text. A useful feature of the local approach is that the conditions for the robustness of the estimation and coverage properties do not depend on the way the consistent critical values are generated (e.g., bootstrap or other means). The conditions are simple to check and apply to *any* consistent method of estimating a critical value.

KEYWORDS: Contiguous perturbation, uniformity.

ROBUSTNESS TO CONTIGUOUS PERTURBATIONS OF P

THE IDEA OF FOCUSING on the local perturbations follows its uses in the confidence interval literature; see notably Dufour (1997) and Pötscher (1991). Intuitively, contiguous perturbations of P cannot be statistically detected with certainty, and we therefore want to make sure that contiguous changes in P do not affect the coverage properties of confidence regions. An alternative motivation is that, in the asymptotic context, the relevant parameter space for nuisance parameters consists of contiguous parameter values, which is a standard approach in asymptotic efficiency analysis; see van der Vaart (1998, Chap. 8.7). In fact, minimal coverage under contiguous sequences implies local uniform coverage, when the local nuisance parameters are allowed to vary over a compact set.

We focus on examining the robustness of the main estimation and inferential results, the ones stated in Theorems 3.1 and 3.3.

S.1. *Regular Cases*

Consider a triangular sequence of probability measures $\{P_{n,\gamma}, n = 1, 2, \ldots\}$, where γ is an index of a sequence in Γ and $\{P_{n,\gamma}, \gamma \in \Gamma, n = 1, \ldots\} \subseteq \mathcal{P}$. Let $P^n_{n,\gamma}$ denote the law of data $w_1, \ldots, w_n$ under $P_{n,\gamma}$. Each $\gamma \in \Gamma$ is such that $P^n_{n,\gamma}$ is contiguous to P^n, the law of data $w_1, \ldots, w_n$ under P, namely $P^n(A_n) = o(1)$ implies $P^n_{n,\gamma}(A_n) = o(1)$ for any sequence of measurable events A_n.[1] In what follows, notation $\Theta_I(P)$ is used to reflect that the identification region Θ_I depends on P, the law of the data. Similarly, notation $c(\alpha, P)$ is used to denote the dependence of the α-quantile of $\mathcal{C}$ on P.

[1]Throughout this supplement, measurable events A_n are events that are measurable with respect to $(\Omega, \mathcal{F})$ completed with respect to both P^n and $P^n_{n,\gamma}$.

LEMMA S.1—Conditions for Maintaining Consistency, Rates of Convergence, and Coverage: (1) *Assume that Conditions* C.1 *and* C.2 *hold*, *with* $\{P_{n,\gamma}\}$ *replacing* $\{P\}$ *for each* $\gamma \in \Gamma$. *Then the conclusions of Theorem* 3.1 *hold*. (2) *Assume that Conditions* C.1, C.2, *and* C.4 *hold under* $\{P_{n,\gamma}\}$ *in place of* $\{P\}$ *for any* $\gamma \in \Gamma$, *with the common limit random variable* $\mathcal{C}$, *the distribution of which does not depend on* γ. *Consider any consistent critical value* $\widehat{c} \to_p c(\alpha, P)$ *under* $\{P\}$. *Then for each* $\gamma \in \Gamma$, $\liminf_{n\to\infty} P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq \alpha$ *and* $= \alpha$ *if* $P\{\mathcal{C} > 0\} \geq \alpha$.

The first result states that consistency and rates of convergence will be preserved under sequences as long as Conditions C.1 and C.2 hold under sequences (replacing P with $P_{n,\gamma}$ and Θ_I with $\Theta_I(P_{n,\gamma})$ should cause no ambiguity in the restatement of Conditions C.1 and C.2). The second result of the lemma addresses coverage properties in the *regular case*—when the limit of $\mathcal{C}_n$ does *not* depend on the local sequence. The definition of regularity follows that given by van der Vaart and Wellner (1996, p. 413). Note that the coverage result is *independent* of the way the critical value is estimated.

Conditions of Lemma S.1 are verified in our principal applications as follows:

CONDITION M.3—Moment Equalities: *Suppose that Condition* M.1 *holds for each* $P \in \mathcal{P}$ *and that* (a) *the partial identification condition* (4.1) *holds uniformly in* $\mathcal{P}$, (b) $G(\theta) = \lim_n \nabla_\theta E_{P_{\gamma,n}}[m_i(\theta)]$ *exists and is continuous over a neighborhood of* Θ *for each* $\gamma \in \Gamma$, (c) *the Donsker condition* (4.2) *holds under* $\{P_{n,\gamma}\}$ *in place of* $\{P\}$ *for each* $\gamma \in \Gamma$, *with the common limit Gaussian process* $\Delta(\theta)$, (d) $E_{P_{n,\gamma}}[m_i(\theta)] = E_P[m_i(\theta)] + o(1)$ *for each* $\gamma \in \Gamma$, *and* (e) $d_H(\Theta_I(P_{n,\gamma}), \Theta_I(P)) = o(1)$ *for each* $\gamma \in \Gamma$.

CONDITION M.4—Moment Inequalities: *Suppose that Condition* M.1 *holds for each* $P \in \mathcal{P}$ *and that* (a) *the partial identification condition* (4.5) *holds uniformly in* $\mathcal{P}$, (b) $G(\theta) = \lim_n \nabla_\theta E_{P_{\gamma,n}}[m_i(\theta)]$ *exists and is continuous over a neighborhood of* Θ *for each* $\gamma \in \Gamma$, (c) *the Donsker condition* (4.2) *holds under* $\{P_{n,\gamma}\}$ *in place of* $\{P\}$ *for each* $\gamma \in \Gamma$, *with the common limit Gaussian process* $\Delta(\theta)$, (d) $E_{P_{n,\gamma}}[m_i(\theta)] = E_P[m_i(\theta)] + o(1)$ *for each* $\gamma \in \Gamma$, *and* (e) $d_H(\Theta_I(P_{n,\gamma}), \Theta_I(P)) = o(1)$ *and* $d_H(\Theta_{\mathcal{J}}(P_{n,\gamma}), \Theta_{\mathcal{J}}(P)) = o(1)$ *for each* $\mathcal{J}$ *and each* $\gamma \in \Gamma$.

Conditions M.3(a) and M.4(a) impose a locally uniform partial identifiability. Sufficient conditions for conditions (c) are well known and are given in van der Vaart and Wellner (1996, p. 173), including a quadratic-mean-differentiability condition (van der Vaart and Wellner (1996, p. 406)). The principal conditions are Conditions M.3(e) and M.4(e), which require that the perturbations of P affect the identification region smoothly.

LEMMA S.2—Coverage, Consistency, and Rates under Regular Sequences in Moment Condition Models: (1) *Condition* M.3 *implies conditions of Lemma* A.1. (2) *Condition* M.4 *implies conditions of Lemma* A.1.

EXAMPLE 1—Continued: It is helpful to illustrate Condition M.4(a)–(e) via a simple example. Recall the example of interval-censored Y without covariates, in which case $\Theta_I(P) = [E_P[Y_1], E_P[Y_2]]$ and suppose $Y_1 \leq Y_2$ P-almost surely for all $P \in \mathcal{P}$ and that (Y_1, Y_2) are uniformly Donsker in $\mathcal{P}$. Conditions for the Donskerness uniformly in $\mathcal{P}$ are well known; see van der Vaart and Wellner (1996, pp. 168–170). Then Condition M.4(a)–(d) easily follows. To verify Condition M.4(e), note that by contiguity and uniform integrability implied by the uniform-in-$\mathcal{P}$ Donskerness, $(E_{P_{n,\gamma}}[Y_1], E_{P_{n,\gamma}}[Y_2]) \to (E_P[Y_1], E_P[Y_2])$, including the case of $[E_P[Y_1], E_P[Y_2]]$ being a singleton.

Conditions M.3 and M.4 are reasonable in many examples we have considered, provided the boundary of Θ_I (in $\mathbb{R}^d$) is strongly identified. Conditions M.3 and M.4 are not expected to hold otherwise. Therefore, the models with weak identification (cf. Dufour (1997)) that are local to nonidentification are not covered by the framework of regular sequences. This motivates the analysis in the next section.

S.2. *Nonregular Cases*

Consider the case where $\mathcal{C}_n$ is *nonregular*. That is, the limit distribution of $\mathcal{C}_n$ under the local sequence $\{P_{n,\gamma}\}$ depends on γ. In this case, the coverage under local sequences depends on whether the distribution of $\mathcal{C}_n$ under local sequences is stochastically dominated in large samples by the distribution under fixed sequence $\{P\}$. The following lemma addresses nonregular cases, showing that the main results will be preserved in a greater generality.

LEMMA S.3—Maintaining Partial Consistency and Minimal Coverage Under Nonregular Sequences: (1) *Suppose that* $\sup_{\Theta_I(P_{n,\gamma})} Q_n = O_{p_{n,\gamma}}(1/a_n)$ *under* $\{P_{n,\gamma}\}$. *Then* $\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})$ *wp* $\to 1$, *provided* $\widehat{c} \to_p \infty$, *under* $\{P_{n,\gamma}\}$. (2) *Let there be any estimate* $\widehat{c} \to_p c(\alpha, P)$ *under* $\{P\}$. *Suppose that Condition* C.4 *holds under the fixed* P *with the limit real variable* $\mathcal{C}$ *that has* α*-quantile* $c(\alpha, P)$. *Suppose that for each* $\gamma \in \Gamma$ *and any sequence* $\epsilon_n \downarrow 0$, *we have*

$$\liminf_{n\to\infty} P_{n,\gamma}\big[\mathcal{C}_n \leq (c(\alpha, P) - \epsilon_n) \vee 0\big] \geq \alpha. \tag{S.1}$$

Consider any estimate $\widehat{c} \to_p c(\alpha, P)$ *under* $\{P\}$, *for instance, that was provided in Section* 3 *or* 4. *Then for each* $\gamma \in \Gamma$,

$$\liminf_{n\to\infty} P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq \alpha. \tag{S.2}$$

Note again that the result is independent of the way the critical value is estimated.

EXAMPLE 4—Weak Instrumental Variable (IV): The point of this lemma can be illustrated using a very simple IV example with one regressor:

$$\begin{aligned} &Y=\theta_0 X+\epsilon, \quad \theta_0\in\Theta \text{ (compact) } \subset\mathbb{R}, \quad X=0\cdot Z+v, \\ &(\epsilon,v)|Z\sim N(0,\Omega), \quad Z\sim N(\mu,\sigma_Z^2). \end{aligned} \tag{S.3}$$

The identification region is $\Theta_I(P)=\Theta$, that is, we have complete nonidentification. Assume that independent and identically distributed sampling and other conditions as in Section 4 hold under P. Now consider a sequence of models where

$$\begin{aligned} &Y=\theta_0 X+\epsilon, \quad X=(\rho/\sqrt{n})Z+v, \\ &(\epsilon,v)|Z\sim N(0,\Omega), \quad Z\sim N(0,\sigma_Z^2). \end{aligned} \tag{S.4}$$

Let $\gamma=\{\rho\}$. Let $P^n_{n,\gamma}$ denote the law of vector $(Y_i,X_i,Z_i,i\le n)$ in (S.4); it is contiguous to the law P^n. Let $P_{n,\gamma}$ denote the law of the infinite independent and identically distributed sequence $(Y_i,X_i,Z_i,i<\infty)$ generated according to (S.4). Note that $\Theta_I(P_{n,\gamma})=\Theta_I(P)=\Theta$ if $\rho=0$ and $\Theta_I(P_{n,\gamma})=\theta_0\in\Theta_I(P)$ if $\rho\ne 0$. This implies that the weak limit of $\mathcal{C}_n$ under $P_{n,\gamma}$ with $\rho\ne 0$ is stochastically smaller than the weak limit of $\mathcal{C}_n$ under $P_{n,\gamma}$ with $\rho=0$, since

$$\sup_{\theta_0}\|\Delta(\theta)'W^{1/2}(\theta)\|^2\le\sup_{\Theta}\|\Delta(\theta)'W^{1/2}(\theta)\|^2. \tag{S.5}$$

Therefore, the α-quantile of the right side is bigger than the α-quantile of the left side, so (S.1) is satisfied. Therefore, for each $\rho\in\mathbb{R}^d$ and $\gamma=\{\rho\}$, $\liminf_{n\to\infty}P_{n,\gamma}\{\Theta_I(P_{n,\gamma})\subseteq C_n(\widehat{c})\}\ge\alpha$.

Next we consider more general local parameter sequences $\gamma=\{\rho_n\}$ with $\rho_n\in K$ for each n, where K is a compact subset of $\mathbb{R}$. Let Γ denote the set of all these sequences. The limit under each convergent subsequence $\rho_n\to\rho$ is either the left or the right side of (S.5). Hence, for each sequence $\{\gamma\}$ in Γ and each sequence $\epsilon_n\searrow 0$,

$$\begin{aligned} &\liminf_{n\to\infty}P_{n,\gamma}\Big[\sup_{\Theta_I(P_{n,\gamma})}\|\Delta(\theta)'W^{1/2}(\theta)\|^2\le(c(\alpha,P)-\epsilon_n)\vee 0\Big] \\ &\quad\ge\liminf_{n\to\infty}P\Big[\sup_{\Theta}\|\Delta(\theta)'W^{1/2}(\theta)\|^2\le(c(\alpha,P)-\epsilon_n)\vee 0\Big]\ge\alpha. \end{aligned} \tag{S.6}$$

This implies by Lemma S.3 that $\liminf_{n\to\infty}P_{n,\gamma}\{\Theta_I(P_{n,\gamma})\subseteq C_n(\widehat{c})\}\ge\alpha$. Equivalently, for K denoting any nonempty compact subset of $\mathbb{R}$,

$$\inf_K\liminf_{n\to\infty}\inf_{\rho\in K}P_{n,\rho}\{\Theta_I(P_{n,\rho})\subseteq C_n(\widehat{c})\}\ge\alpha, \tag{S.7}$$

where $P_{n,\rho}$ corresponds to the law of the model (S.4). This coverage property is in the spirit of local asymptotic minimax analysis of estimation; see van der Vaart (1998, Chap. 8.7).

S.3. *Proof of Lemma S.1*

PROOF OF PART (1): The proof is straightforward by substituting $\{P_{n,\gamma}\}$ in place of the fixed sequence P in the proof of Theorem 3.1. *Q.E.D.*

PROOF OF PART (2): We have that $\widehat{c} \to_p c(\alpha, P)$ under $\{P\}$. By contiguity, $\widehat{c} \to_p c(\alpha, P)$ under $\{P_{n,\gamma}\}$. Therefore, $P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq P_{n,\gamma}[\mathcal{C}_n \leq \widehat{c}] = P_{n,\gamma}[\mathcal{C}_n \leq c(\alpha, P) + o_{p_{n,\gamma}}(1)] = P[\mathcal{C} \leq c(\alpha, P)] + o(1)$ by assumption that $P_{n,\gamma}[\mathcal{C}_n \leq c] \to P[\mathcal{C} \leq c]$ for all $c \geq 0$, by $\widehat{c} \geq 0$, and by continuity of the distribution function $c \mapsto P[\mathcal{C} \leq c]$ on $[0, \infty)$. *Q.E.D.*

S.4. *Proof of Lemma S.2*

PROOF OF PART (1): The proof is straightforward by repeating Steps 1–4 in the proof of Theorem 4.1, having replaced P with $P_{n,\gamma}$, Θ_I with $\Theta_I(P_{n,\gamma})$, $o_p(1)$ with $o_{p_{n,\gamma}}(1)$, and so forth, and then noting that $\sup_{\Theta_I(P_{n,\gamma})} \|\Delta(\theta)' W^{1/2}(\theta)\|^2 = \sup_{\Theta_I(P)} \|\Delta(\theta)' W^{1/2}(\theta)\|^2 + o_{p_{n,\gamma}}(1)$ by equicontinuity of $\theta \mapsto \Delta(\theta)' W^{1/2}(\theta)$ and by $d_H(\Theta_I(P_{n,\gamma}), \Theta_I(P)) = o(1)$ imposed in Condition M.3(e). By Condition M.3(b), $\Delta(\theta)$ does not depend on γ, and by contiguity, $W(\theta)$ does not either. Hence the limit variable $\mathcal{C} := \sup_{\Theta_I(P)} \|\Delta(\theta)' W^{1/2}(\theta)\|^2$ does not depend on γ. *Q.E.D.*

PROOF OF PART (2): The proof is straightforward by repeating Steps 1–4 in the proof of Theorem 4.1, having replaced P with $P_{n,\gamma}$, Θ_I with $\Theta_I(P_{n,\gamma})$, and $o_p(1)$ with $o_{p_{n,\gamma}}(1)$. The exception is that in Step 2, we need to define $\xi(\theta) = \lim_n \sqrt{n} E_P[m_i(\theta)]$ under fixed sequence $\{P\}$. Note that the key inequality (A.5) in Lemma A.1 is preserved under sequences $\{P_{n,\gamma}\}$. In the proof of Lemma A.1, the convergent subsequence $\{\theta_n\}$ in $\Theta_I(P)$ is replaced by the convergent subsequence $\{\theta_n\}$ in $\Theta_I(P_{n,\gamma})$, where convergent means $\theta_n \to \theta \in \Theta_I(P)$. Since we care only about $\mathcal{C}_n$ in this lemma, in repeating the proof of Lemma A.1, we set $\lambda = 0$ and only consider the set $V_n^0 = V_\infty^0 = \Theta_I(P_{n,\gamma}) \times \{0\}$. In addition, we note that for every $\mathcal{J}$, $d_H(\Theta_{\mathcal{J}}(P_{n,\gamma}), \Theta_{\mathcal{J}}(P)) = o(1)$ by Condition M.4(e), so that

$$\max_{\mathcal{J}} \sup_{\Theta_{\mathcal{J}}(P_{n,\gamma})} \sum_{j \in \mathcal{J}} \left|(\Delta_j(\theta) + G_j(\theta)'\lambda) W_{jj}^{1/2}(\theta) + o_{p_{n,\gamma}}(1)\right|_+^2$$

$$= \max_{\mathcal{J}} \sup_{\Theta_{\mathcal{J}}(P)} \sum_{j \in \mathcal{J}} \left|(\Delta_j(\theta) + G_j(\theta)'\lambda) W_{jj}^{1/2}(\theta) + o_{p_{n,\gamma}}(1)\right|_+^2.$$

Here we utilized equicontinuity of $\theta \mapsto \Delta(\theta)' W^{1/2}(\theta)$ and the independence of $\Delta(\theta)$ and $W(\theta)$ from γ by Condition M.4(b) and contiguity, respectively. The result of the modified Step 2 can be stated then as

$$\sup_{\Theta_I(P_{n,\gamma})} \ell_n(\theta, 0) =_d \max_{\mathcal{J}} \sup_{\theta \in \Theta_{\mathcal{J}}(P)} \sum_{j \in \mathcal{J}} \left|(\Delta_j(\theta)) W_{jj}^{1/2}(\theta) + o_{p_{n,\gamma}}(1)\right|_+^2.$$

Hence

$$\mathcal{C} = \max_{\mathcal{J}} \sup_{\theta \in \Theta_{\mathcal{J}}(P)} \sum_{j \in \mathcal{J}} \left|(\Delta_j(\theta)) W_{jj}^{1/2}(\theta)\right|_+^2,$$

which does not depend on γ. *Q.E.D.*

S.5. *Proof of Lemma S.3*

PROOF OF PART (1): Under $\{P_{n,\gamma}\}$, wp $\to 1$, by construction of $\widehat{c}$, $\sup_{\Theta_I(P_{n,\gamma})} Q_n = O_{p_{n,\gamma}}(1/a_n) < \widehat{c}/a_n$, which implies $\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})$. *Q.E.D.*

v

PROOF OF PART (2): We have that $\widehat{c} \to_p c(\alpha, P)$ under $\{P\}$. By contiguity, $\widehat{c} \to_p c(\alpha, P)$ under $\{P_{n,\gamma}\}$. Hence $P_{n,\gamma}\{\Theta_I(P_{n,\gamma}) \subseteq C_n(\widehat{c})\} \geq P_{n,\gamma}\{\mathcal{C}_n \leq \widehat{c}\} \geq P_{n,\gamma}\{\mathcal{C}_n \leq (c(\alpha, P) - \epsilon_n) \vee 0\}$ for some $\epsilon_n \downarrow 0$. The conclusion follows from the assumption that $\liminf_{n\to\infty} P_{n,\gamma}\{\mathcal{C}_n \leq (c(\alpha, P) - \epsilon_n) \vee 0\} \geq \alpha$. *Q.E.D.*

Dept. of Economics and Operations Research Center, MIT, 50 Memorial Drive, Cambridge, MA 02139, U.S.A.; chern@mit.edu,

Dept. of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305-6072, U.S.A.; han.hong@duke.edu,

and

Dept. of Economics, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208, U.S.A.; tamer@northwest.edu.

Manuscript received April, 2004; final revision received March, 2007.

REFERENCES

DUFOUR, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica*, 65, 1365–1387. [1,3]

PÖTSCHER, B. M. (1991): "Effects of Model Selection on Inference," *Econometric Theory*, 7, 163–185. [1]

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge, U.K.: Cambridge University Press. [1,5]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer-Verlag. [2,3]