Econometrica Supplementary Material

SUPPLEMENT TO "AVERAGE AND QUANTILE EFFECTS IN NONSEPARABLE PANEL MODELS"

(*Econometrica*, Vol. 81, No. 2, March 2013, 535–580)

BY VICTOR CHERNOZHUKOV, IVÁN FERNÁNDEZ-VAL, JINYONG HAHN, AND WHITNEY NEWEY

S.1. INTRODUCTION

IN THIS SUPPLEMENTAL MATERIAL, we provide omitted discussions, results, and proofs by section in the same order they are referred to in the paper. Let w.p.a.1 denote "with probability approaching 1" and C denote a generic constant that may be different in different uses.

S.2. SUPPLEMENTS TO SECTION 2

We begin with the omitted discussion and results referred to in Section 2 of the paper. These concern the general, nonseparable model of Assumptions 1–3 and apply whether or not the regressors are discrete.

S.2.1. *Time Homogeneity in the Linear Model*

We first show that Assumption 2 is a natural generalization of the following linear model:

$$Y_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it}, \quad E[X_{is}\varepsilon_{it}] = 0 \quad \text{for all } s \text{ and } t. \tag{S.1}$$

This is a standard linear model that leads to consistency of the within and other estimators. Let $\overline{E}(\cdot|X_i)$ denote the linear projection on $\text{vec}(X_i)$, as in Chamberlain (1982).

THEOREM A1: *Suppose that Y_i and X_i have finite second moments. Then equation* (S.1) *is satisfied if and only if there is $\tilde{\varepsilon}_{it}$ with*

$$Y_{it} = X_{it}'\beta_0 + \tilde{\varepsilon}_{it}, \quad \overline{E}(\tilde{\varepsilon}_{it}|X_i) = \overline{E}(\tilde{\varepsilon}_{i1}|X_i) \qquad (t = 2, \dots, T). \tag{S.2}$$

PROOF: If equation (S.1) is satisfied, let $\tilde{\varepsilon}_{it} = \alpha_i + \varepsilon_{it}$. By orthogonality of ε_{it}, with X_{is} for all s and t, we have $\overline{E}(\varepsilon_{it}|X_i) = 0$ for all t, so that

$$\begin{aligned}\overline{E}(\tilde{\varepsilon}_{it}|X_i) &= \overline{E}(\alpha_i|X_i) + \overline{E}(\varepsilon_{it}|X_i) = \overline{E}(\alpha_i|X_i)\\ &= \overline{E}(\alpha_i|X_i) + \overline{E}(\varepsilon_{i1}|X_i) = \overline{E}(\tilde{\varepsilon}_{i1}|X_i).\end{aligned}$$

Now suppose equation (S.2) is satisfied. Let $\alpha_i = \overline{E}[\tilde{\varepsilon}_{i1}|X_i]$ and $\varepsilon_{it} = \tilde{\varepsilon}_{it} - \alpha_i$. Then $Y_{it} = X_{it}'\beta_0 + \alpha_i + \varepsilon_{it}$ by construction and

$$E[X_{is}\varepsilon_{it}] = E\big[X_{is}\big(\tilde{\varepsilon}_{it} - \overline{E}[\tilde{\varepsilon}_{i1}|X_i]\big)\big] = E\big[X_{is}\big(\tilde{\varepsilon}_{it} - \overline{E}[\tilde{\varepsilon}_{it}|X_i]\big)\big] = 0,$$

 DOI: 10.3982/ECTA8405

where the second equality follows by $\overline{E}(\tilde{\varepsilon}_{it}|X_i) = \overline{E}(\tilde{\varepsilon}_{i1}|X_i)$ and the third equality by orthogonality of each element of X_i with the projection residual. *Q.E.D.*

This result shows that the standard linear model of equation (S.1) is equivalent to the model of equation (S.2). The second model is one that satisfies a time-homogeneity condition analogous to Assumption 2. In equation (S.2), the linear projection of the disturbance on the elements of X_i is time invariant. What Assumption 2 does is strengthen this to time invariance of the conditional distribution. This strengthening seems like a natural thing to do when moving from a linear model to a nonlinear, nonseparable model.

S.2.2. *Relationship Between Static and Dynamic Models*

We next show that the static model is nested within the dynamic model.

THEOREM A2: *If Assumptions* 1 *and* 2 *are satisfied*, *then Assumptions* 1 *and* 3 *are satisfied*.

PROOF: Note that Assumptions 1 and 2 allow some flexibility in the definition of α_i, because Assumption 1 just specifies that there exists α_i with $Y_{it} = g_0(X_{it}, \alpha_i, \varepsilon_{it})$. This equation continues to hold if more variables are added to α_i. Furthermore, we can add any function of X_i to α_i without changing Assumption 2. Let $\tilde{\alpha}_i = (\alpha_i, X_i)$. Then Assumptions 1 and 2 are also satisfied for this $\tilde{\alpha}_i$. Furthermore, since $X_{it}, \dots, X_{i1}$ are included in $\tilde{\alpha}$ and Assumption 2 for the original α_i implies that $\varepsilon_{it}|\tilde{\alpha}_i \overset{d}{=} \varepsilon_{i1}|\tilde{\alpha}_i$, we have

$$\varepsilon_{it}|X_{it}, \dots, X_{i1}, \tilde{\alpha}_i \overset{d}{=} \varepsilon_{it}|\tilde{\alpha}_i \overset{d}{=} \varepsilon_{i1}|\tilde{\alpha}_i \overset{d}{=} \varepsilon_{i1}|X_{i1}, \tilde{\alpha}_i.$$

Thus we see that Assumptions 1 and 2 imply existence of $\alpha_i = \tilde{\alpha}_i$ such that Assumptions 1 and 3 are also satisfied. That is, Assumptions 1 and 2 imply Assumptions 1 and 3. *Q.E.D.*

S.2.3. *Relationship Between Nonseparable Models and Conditional-Mean Models*

Next we show that the nonseparable models given here imply conditional-mean models where the ATE is also the conditional-mean ATE.

THEOREM A3: *Suppose that Assumption* 1 *is satisfied and* $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ *for all* x. *If Assumption* 2 *is satisfied*, *then*, *for* $\tilde{\alpha}_i = X_i$ *and* $m_0(x, \tilde{\alpha}) = \int g_0(x, \alpha, \varepsilon)\, dF(\alpha, \varepsilon|\tilde{\alpha})$,

$$E[Y_{it}|X_i, \tilde{\alpha}_i] = m_0(X_{it}, \tilde{\alpha}_i), \quad \mu(x) = \int m_0(x, \tilde{\alpha})\, dF(\tilde{\alpha}).$$

If Assumption 3 *is satisfied*, *then*, *for* $\tilde{\alpha} = (\alpha, X_1)$ *and* $m_0(x, \tilde{\alpha}) = \int g_0(x, \alpha, \varepsilon)\, dF(\varepsilon|\tilde{\alpha})$,

$$E[Y_{it}|X_{it}, \dots, X_{i1}, \tilde{\alpha}_i] = m_0(X_{it}, \tilde{\alpha}_i), \quad \mu(x) = \int m_0(x, \tilde{\alpha}) F(d\tilde{\alpha}).$$

PROOF: By Assumption 2, for $\tilde{\alpha} = X$ and $m_0(x, \tilde{\alpha}) = \int g_0(x, \alpha, \varepsilon)\, dF(\alpha, \varepsilon|X)$ we have

$$\begin{aligned} E[Y_{it}|X_i, \tilde{\alpha}_i] &= E\big[g_0(X_{it}, \alpha_i, \varepsilon_{it})|X_i\big] \\ &= \int g_0(X_{it}, \alpha, \varepsilon)\, dF(\alpha, \varepsilon|\tilde{\alpha}_i) = m_0(X_{it}, \tilde{\alpha}_i), \end{aligned}$$

$$\int m_0(x, \tilde{\alpha})\, dF(\tilde{\alpha}) = \int g_0(x, \alpha, \varepsilon)\, dF(\alpha, \varepsilon|\tilde{\alpha})\, dF(\tilde{\alpha}) = \mu(x).$$

Similarly, Assumption 3 implies, for $\tilde{\alpha}_i = (\alpha_i, X_{1i})$,

$$\begin{aligned} E[Y_{it}|X_{it}, \dots, X_{i1}, \tilde{\alpha}_i] &= \int g_0(X_{it}, \alpha_i, \varepsilon)\, dF(\varepsilon|X_{it}, \dots, X_{i1}, \alpha_i) \\ &= \int g_0(X_{it}, \alpha_i, \varepsilon)\, dF(\varepsilon|\alpha_i, X_{i1}) = m_0(X_{it}, \tilde{\alpha}_i), \end{aligned}$$

$$\begin{aligned} \int m_0(x, \tilde{\alpha})\, dF(\tilde{\alpha}) &= \int g_0(x, \alpha, \varepsilon)\, dF(\varepsilon|\alpha, X_1)\, dF(\alpha, X_1) \\ &= \int g_0(x, \alpha, \varepsilon)\, dF(\varepsilon, \alpha, X_1) = \mu(x). \end{aligned}$$

Q.E.D.

It may be helpful to explain this result and relate it to Chamberlain (1982). First, it should be noted that Assumptions 1 and 2 only assume the existence of some α_i such that the conditions are satisfied. Thus, we are free to choose α_i in whatever way is convenient. A convenient choice for Theorem A3 turns out to be $\tilde{\alpha}_i = X_i$, where we use the $\tilde{\alpha}_i$ notation to distinguish this time-invariant effect from the one in Assumptions 1 and 2. Note then that the first conclusion implies that, for $m_0(x, X) = \int g(x, \alpha, \varepsilon)\, dF(\alpha, \varepsilon|X)$,

$$E[Y_{it}|X_i] = m_0(X_{it}, X_i). \tag{S.3}$$

This statement has no content for any one time period, because the effect of X_{it} in the first argument of $m(X_{it}, X_i)$ is indistinguishable from the effect of X_{it} that appears in the second argument. However, for multiple time periods it does have content, because $m_0(x, X)$ is time invariant. Equation (S.3) implies that the effect of changing X_{it} on $E[Y_{it}|X_i]$ will be different than the effect on $E[Y_{is}|X_i]$ for $s \neq t$. Furthermore, this form leads directly to identification of conditional-mean ATE conditioned on X_i. For any X_i where $X_{it} = x^b$ and $X_{is} = x^a$ for some t and s,

$$E[Y_{is} - Y_{it}|X_i] = m_0(x^a, X_i) - m_0(x^b, X_i),$$

that is, a conditional-mean ATE given X_i.

It may also help to think of $m(X_{it}, X_i)$ as a nonlinear version of Chamberlain's (1982) multivariate regression for panel data. In the linear model of equation (S.1), for $\overline{E}[\alpha_i|X_i] = \pi' \operatorname{vec}(X_i)$, we have

$$\overline{E}[Y_{it}|X_i] = X_{it}'\beta_0 + \pi' \operatorname{vec}(X_i) = \overline{m}(X_{it}, X_i),$$
$$\overline{m}(x, X) = x'\beta_0 + \pi' \operatorname{vec}(X).$$

For a single time period, β_0 is indistinguishable from coefficients in π, but multiple time periods can be used to identify β_0 from these regressions. Equation (S.3) is like this except it is jointly nonlinear in its first and second arguments.

S.3. SUPPLEMENTS TO SECTION 3

S.3.1. *Auxiliary Results*

We turn now to identification and estimation with discrete regressors in the static case. Here we use the idea that "time is an instrument" or "time is randomly assigned." This allows us to vary the time period so as to match x with X_{it} and achieve identification.

The following lemma applies this idea to obtain specific results. Let $g_{it}(x) = g_0(x, \alpha_i, \varepsilon_{it})$.

LEMMA A4: *If Assumptions* 1 *and* 2 *are satisfied*, *then*

$$E[\overline{G}_i(y, x)|X_i] = 1(T_i(x) > 0)E\left[\Phi\left(\frac{y - g_{i1}(x)}{h}\right)|X_i\right].$$

If, *in addition*, $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ *for all* x, *then*

$$E[\overline{Y}_i(x)|X_i] = 1(T_i(x) > 0)E[g_{i1}(x)|X_i].$$

PROOF: By Assumptions 1 and 2,

$$E\left[1(X_{it} = x)\Phi\left(\frac{y - Y_{it}}{h}\right)|X_i\right]$$
$$= E\left[1(X_{it} = x)\Phi\left(\frac{y - g_{it}(x)}{h}\right)|X_i\right]$$

$$= 1(X_{it} = x)E\left[\Phi\left(\frac{y - g_{it}(x)}{h}\right)|X_i\right]$$
$$= 1(X_{it} = x)E\left[\Phi\left(\frac{y - g_{i1}(x)}{h}\right)|X_i\right].$$

Therefore, we have

$$E\big[\overline{G}_i(y, x)|X_i\big]$$
$$= 1\big(T_i(x) > 0\big)T_i(x)^{-1}\sum_{t=1}^{T} E\left[1(X_{it} = x)\Phi\left(\frac{y - Y_{it}}{h}\right)|X_i\right]$$
$$= 1\big(T_i(x) > 0\big)T_i(x)^{-1}\sum_{t=1}^{T} 1(X_{it} = x)E\left[\Phi\left(\frac{y - g_{i1}(x)}{h}\right)|X_i\right]$$
$$= 1\big(T_i(x) > 0\big)E\left[\Phi\left(\frac{y - g_{i1}(x)}{h}\right)|X_i\right].$$

We also have

$$E\big[1(X_{it} = x)Y_{it}|X_i\big]$$
$$= E\big[1(X_{it} = x)g_{it}(x)|X_i\big] = 1(X_{it} = x)E\big[g_{it}(x)|X_i\big]$$
$$= 1(X_{it} = x)E\big[g_{i1}(x)|X_i\big],$$

so the second conclusion follows similarly to the first. *Q.E.D.*

We can use the previous result to show how δ is identified.

LEMMA A5: *If Assumptions* 1 *and* 2 *are satisfied*, $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ *for all* x, *and* $\Pr(D_i = 1) > 0$, *then* $\delta = E[D_i\{\overline{Y}_i(x^a) - \overline{Y}_i(x^b)\}]/E[D_i]$.

PROOF: Note that $D_i = D_i 1(T_i(x^b) > 0) = D_i 1(T_i(x^a) > 0)$. Therefore, by Lemma A4,

$$E\big[D_i\big\{\overline{Y}_i\big(x^a\big) - \overline{Y}_i\big(x^b\big)\big\}|X_i\big]$$
$$= D_i E\big[\overline{Y}_i\big(x^a\big)|X_i\big] - D_i E\big[\overline{Y}_i\big(x^b\big)|X_i\big]$$
$$= D_i 1\big(T_i\big(x^a\big) > 0\big)E\big[g_{i1}\big(x^a\big)|X_i\big] - D_i 1\big(T_i\big(x^b\big) > 0\big)E\big[g_{i1}\big(x^b\big)|X_i\big]$$
$$= D_i E\big[g_{i1}\big(x^a\big) - g_{i1}\big(x^b\big)|X_i\big] = E\big[D_i\big\{g_{i1}\big(x^a\big) - g_{i1}\big(x^b\big)\big\}|X_i\big].$$

The conclusion then follows by iterated expectations. *Q.E.D.*

The asymptotic normality of $\hat{\delta}$ and consistency of the asymptotic variance estimator are simple applications of standard theory, as in the following result, that forms a prototype for the asymptotic normality of the nonparametric ATE bounds. Let $P = E[D_i]$.

THEOREM A6: *If Assumptions* 1 *and* 2 *are satisfied*, $E[|g_0(x, \alpha_i, \varepsilon_{it})|^2] < \infty$ *for all* x, *and* $\Pr(D_i = 1) > 0$, *then* $\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} N(0, V)$ *and* $\sum_{i=1}^n \hat{\psi}_i^2/n \xrightarrow{p} V$, *where* $V = E[\psi_i^2]$ *and* $\psi_i = P^{-1} D_i[\overline{Y}_i(x^a) - \overline{Y}_i(x^b) - \delta]$.

PROOF: Let $d_i = D_i\{\overline{Y}_i(x^a) - \overline{Y}_i(x^b)\}$, so that $\hat{\delta} = \overline{d}/\overline{D}$. By the central limit theorem (CLT), $\overline{d}$ and $\overline{D}$ are root-n consistent for $\mu_d = E[d_i]$ and P. Then, by $P > 0$ and $\delta = \mu_d/P$,

$$\begin{aligned}
\sqrt{n}(\hat{\delta} - \delta) &= \sqrt{n}\left(\frac{\overline{d}}{\overline{D}} - \frac{\mu_d}{P}\right) = \sqrt{n}\overline{D}^{-1}\left[\overline{d} - \mu_d - \delta(\overline{D} - P)\right] \\
&= \sqrt{n}P^{-1}\left[\overline{d} - \mu_d - \delta(\overline{D} - P)\right] + o_p(1) \\
&= \sum_{i=1}^n \psi_i/\sqrt{n} + o_p(1).
\end{aligned}$$

The first conclusion then follows by the CLT. For the second conclusion, note that

$$\begin{aligned}
&\sum_i (\hat{\psi}_i - \psi_i)^2/n \\
&\quad \leq C\left(\overline{D}^{-1} - P^{-1}\right)^2 \sum_i d_i^2/n + C\left(\overline{D}^{-1}\hat{\delta} - P^{-1}\delta\right)^2 \sum_i D_i^2/n \xrightarrow{p} 0.
\end{aligned}$$

Therefore, the second conclusion follows by a standard argument. *Q.E.D.*

We now give an intermediate result that is useful for showing asymptotic normality for the estimator of the identified quantile treatment effect. This will also serve as a prototype for the proofs of Theorems 2 and 3 in the body of the paper. Let $\hat{G}_1(y, x) = \hat{G}(y, x | D_i = 1)$, $G_1(y, x) = G(y, x | D_i = 1)$, $G_i(y, x) = 1(T_i(x) > 0) T_i(x)^{-1} \sum_{t=1}^T 1(X_{it} = x) 1(Y_{it} \leq y)$, and $G_1'(y, x) = \partial G_1(y, x)/\partial y$.

LEMMA A7: *If Assumption* 7 *is satisfied with* $G_\ell(y, x)$ *replaced by* $G_1(y, x)$, *then for any* $0 < \lambda < 1$ *and any* x, *there exists* $\hat{q}_\lambda$ *with* $\hat{G}_1(\hat{q}_\lambda, x) = \lambda$ *satisfying*

$$\sqrt{n}(\hat{q}_\lambda - q_\lambda) = -G_1'(q_\lambda, x)^{-1} \frac{1}{\sqrt{n}} P^{-1} \sum_i D_i\left[G_i(q_\lambda, x) - \lambda\right] + o_p(1).$$

PROOF: Note that $\hat{G}_1(y, x)$ is strictly monotonic increasing in y and converges to 0 and 1 as y goes to $-\infty$ and ∞, respectively. Therefore, there is a unique $\hat{q}_\lambda$ such that $\hat{G}_1(\hat{q}_\lambda, x) = \lambda$. Also, by $G_1(y, x)$ strictly monotonic in y, there is a unique q_λ solving $G_1(q_\lambda, x) = \lambda$. By $G_1(y, x)$ strictly monotonic and continuous, it follows that, for all $\varepsilon > 0$ small enough,

$$0 < G_1(q_\lambda - \varepsilon, x) < G_1(q_\lambda, x) = \lambda.$$

By $\hat{G}_1(q_\lambda - \varepsilon, x) \xrightarrow{p} G_1(q_\lambda - \varepsilon, x)$, it follows that w.p.a.1, for all $y \le q_\lambda - \varepsilon$,

$$\hat{G}_1(y, x) \le \hat{G}_1(q_\lambda - \varepsilon, x) < G_1(q_\lambda, x) = \lambda.$$

Thus, it follows that $\hat{q}_\lambda \ge q_\lambda - \varepsilon$ w.p.a.1. Similarly, it follows that $\hat{q}_\lambda \le q_\lambda + \varepsilon$ w.p.a.1. Since ε is arbitrary, we have $\hat{q}_\lambda \xrightarrow{p} q_\lambda$.

Next, note that $G_1(y, x)$ is differentiable in y by Assumption 7, so that $g_{i1}(x)$ is continuously distributed conditional on $D_i = 1$. Thus, $g_{it}(x)$ is also continuously distributed conditional on $D_i = 1$ by Assumption 2. It follows that, as $h \longrightarrow 0$, $\Phi(\frac{y-g_{it}(x)}{h}) \longrightarrow 1(g_{it}(x) \le y)$ with probability 1. By the dominated convergence theorem, this convergence is also in mean square. Recall that

$$G_i(y, x) = \begin{cases} T_i(x)^{-1} \sum_{t=1}^{T} 1(X_{it} = x) 1(Y_{it} \le y), & T_i(x) > 0, \\ 0, & T_i(x) = 0. \end{cases}$$

We have $\overline{G}_i(y, x) \longrightarrow G_i(y, x)$ in mean square, so that

$$\sum_{i=1}^{n} \big[D_i \overline{G}_i(y, x) - D_i G_i(y, x)\big]/n \xrightarrow{p} 0,$$

$$\sum_{i=1}^{n} \big\{D_i \overline{G}_i(y, x) - E\big[D_i \overline{G}_i(y, x)\big] - D_i G_i(y, x) + E\big[D_i G_i(y, x)\big]\big\}/\sqrt{n} \xrightarrow{p} 0.$$

Let $W_i = g_0(x, \alpha_i, \varepsilon_{i1})$ and $f(w)$ and $F(w)$ denote the p.d.f. and CDF of W_i conditional on $D_i = 1$ and $P = E[D_i]$. Note that $\Phi(\frac{y-w}{h})F(w)$ converges to zero as $w \longrightarrow \infty$ and as $w \longrightarrow -\infty$. Therefore, integration by parts gives

$$\begin{aligned} & E\big[\overline{G}_i(y, x) | D_i = 1\big] \\ &= \int \Phi\left(\frac{y-w}{h}\right) f(w)\, dw = h^{-1} \int \phi\left(\frac{y-w}{h}\right) F(w)\, dw \\ &= \int \phi(u) F(y - hu)\, du = F(y) + \big(h^2/2\big) \int \phi(u) F''(y - \overline{h}u) u^2\, du \\ &= F(y) + o\big(h^2\big) = G_1(y, x) + o\big(h^2\big), \end{aligned}$$

where the fifth equality follows by an expansion

$$F(y-hu)=F(y)-F'(y)hu+F''(y-\overline{h}u)h^2u^2/2,$$

and $\overline{h}$ can depend on u. Therefore, it follows by $E[D_iG_i(q_\lambda,x)]=PG_1(q_\lambda,x)=P\lambda$ that

$$\begin{aligned}
&\sum_{i=1}^{n} D_i\big[\overline{G}_i(q_\lambda,x)-\lambda\big]/\sqrt{n}\\
&\quad=\sum_{i=1}^{n}\big\{D_i\overline{G}_i(q_\lambda,x)-E\big[D_i\overline{G}_i(q_\lambda,x)\big]\big\}/\sqrt{n}\\
&\qquad+\sqrt{n}\big\{E\big[D_i\overline{G}_i(q_\lambda,x)\big]-\lambda P\big\}-\lambda\sum_{i=1}^{n}(D_i-P)/\sqrt{n}\\
&\quad=\sum_{i=1}^{n}\big\{D_iG_i(q_\lambda,x)-E\big[D_iG_i(q_\lambda,x)\big]\big\}/\sqrt{n}+o_p(1)\\
&\qquad+O\big(\sqrt{n}h^2\big)-\lambda\sum_{i=1}^{n}(D_i-P)/\sqrt{n}\\
&\quad=\sum_{i=1}^{n}D_i\big[G_i(q_\lambda,x)-\lambda\big]/\sqrt{n}+o_p(1)=O_p(1).
\end{aligned}$$

Next, note that from standard uniform convergence of kernel density results, $\hat{G}_1'(y,x)$ converges uniformly in probability to $G_1'(y,x)$, where the "prime" superscript denotes the partial derivative with respect to y. Therefore, for $\overline{q}_\lambda \xrightarrow{p} q_\lambda$, $\hat{G}_1'(\overline{q}_\lambda,x)\xrightarrow{p} G_1'(q_\lambda,x)>0$, and hence $\hat{G}_1'(\overline{q}_\lambda,x)^{-1}=O_p(1)$. An expansion then gives $\lambda=\hat{G}_1(\hat{q}_\lambda,x)=\hat{G}_1(q_\lambda,x)+\hat{G}_1'(\overline{q}_\lambda,x)(\hat{q}_\lambda-q_\lambda)$. Solving and inverting gives

$$\begin{aligned}
&\sqrt{n}(\hat{q}_\lambda-q_\lambda)\\
&\quad=-\hat{G}_1'(\overline{q}_\lambda,x)^{-1}\sqrt{n}\big[\hat{G}_1(q_\lambda,x)-\lambda\big]\\
&\quad=-\hat{G}_1'(\overline{q}_\lambda,x)^{-1}\left(\sum_{i=1}^{n}D_i/n\right)^{-1}\sum_{i=1}^{n}D_i\big[\overline{G}_i(q_\lambda,x)-\lambda\big]/\sqrt{n}\\
&\quad=-G_1'(q_\lambda,x)^{-1}P^{-1}\sum_{i=1}^{n}D_i\big[G_i(q_\lambda,x)-\lambda\big]/\sqrt{n}+o_p(1).
\end{aligned}$$

Q.E.D.

THEOREM A8: *If Assumptions* 1, 2, *and* 7 *are satisfied and* $E[D_i] > 0$, *then* $\sqrt{n}(\hat{\delta}_\lambda - \delta_\lambda) \xrightarrow{d} N(0, V_\lambda)$ *and* $\sum_{i=1}^n \hat{\psi}_{\lambda i}^2/n \xrightarrow{p} V_\lambda$, *where* $V_\lambda = E[\psi_{\lambda i}^2]$ *and*

$$\psi_{i\lambda} = -\frac{D_i}{P}\left\{\frac{G_i(q^a, x^a) - \lambda}{G_1'(q^a, x^a)} - \frac{G_i(q^b, x^b) - \lambda}{G_1'(q^b, x^b)}\right\}.$$

PROOF: By Lemma A7, we have

$$\sqrt{n}(\hat{\delta}_\lambda - \delta_\lambda) = \sum_{i=1}^n \psi_{i\lambda}/\sqrt{n} + o_p(1).$$

The CLT gives the first conclusion. Next, note that by $\Phi(v)$ having a bounded derivative,

$$\begin{aligned}&\sum_{i=1}^n \bigl[\overline{G}_i(\hat{q}^a, x^a) - \overline{G}_i(q^a, x^a)\bigr]^2/n \\ &\quad \le Ch^{-1}(\hat{q}^a - q^a) = O_p\bigl((h\sqrt{n})^{-1}\bigr) \xrightarrow{p} 0.\end{aligned}$$

Then by mean square convergence of $\overline{G}_i(q^a, x^a)$ to $G_i(q^a, x^a)$ and the triangle inequality, we have $\sum_{i=1}^n [\overline{G}_i(\hat{q}^a, x^a) - G_i(q^a, x^a)]^2/n \xrightarrow{p} 0$. The second conclusion then follows similarly to the proof of Theorem A6. *Q.E.D.*

S.3.2. *Proof of Theorem 1*

Note that $\sigma_i^2 > 0$ if and only if $D_i = 1$, so that

$$\sigma_i^2 = D_i\sigma_i^2, \quad X_{it} - \overline{X}_i = D_i(X_{it} - \overline{X}_i).$$

Furthermore, since X_{it} is a dummy variable, the usual difference in means formula for the slope of a regression on a constant and dummy variable gives

$$D_i\frac{\sum_{t=1}^T (X_{it} - \overline{X}_i)Y_{it}}{\sum_{t=1}^T (X_{it} - \overline{X}_i)^2} = D_i\bigl\{\overline{Y}_i(1) - \overline{Y}_i(0)\bigr\}.$$

Also, by Khintchine's weak law of large numbers (LLN),

$$\begin{aligned}&n^{-1}(T-1)^{-1}\sum_{i=1}^n\sum_{t=1}^T (X_{it} - \overline{X}_i)^2 \\ &\quad = n^{-1}\sum_{i=1}^n \sigma_i^2 \xrightarrow{p} E\bigl[\sigma_i^2\bigr] = E\bigl[D_i\sigma_i^2\bigr].\end{aligned}$$

Furthermore, by LLN,

$$\begin{aligned}
&n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(X_{it}-\overline{X}_i)Y_{it}\\
&\quad = n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}D_i(X_{it}-\overline{X}_i)Y_{it}\\
&\quad = n^{-1}\sum_{i=1}^{n}D_i\sigma_i^2\{\overline{Y}_i(1)-\overline{Y}_i(0)\}\\
&\quad \xrightarrow{p} E\big[D_i\sigma_i^2\{\overline{Y}_i(1)-\overline{Y}_i(0)\}\big].
\end{aligned}$$

The conclusion then follows by the continuous mapping theorem. *Q.E.D.*

S.4. SUPPLEMENTS TO SECTION 4

Here we include the proof of Theorem 2 as well as bounds that impose monotonicity.

S.4.1. *Proof of Theorem 2*

Let

$$\begin{pmatrix} m_{\ell i} \\ m_{ui} \end{pmatrix} = \begin{pmatrix} \overline{Y}_i(x^a)-\overline{Y}_i(x^b)+B_\ell 1(T_i(x^a)=0)-B_u 1(T_i(x^b)=0) \\ \overline{Y}_i(x^a)-\overline{Y}_i(x^b)+B_u 1(T_i(x^a)=0)-B_\ell 1(T_i(x^b)=0) \end{pmatrix}.$$

Note that $\hat{\Delta}_\ell = \sum_{i=1}^n m_{\ell i}/n$ and $\hat{\Delta}_u = \sum_{i=1}^n m_{ui}/n$. Then, for $\Sigma = \mathrm{Var}((m_{\ell i}, m_{ui}))$, $\Delta_\ell = E[m_{\ell i}]$, and $\Delta_u = E[m_{ui}]$, the first and second conclusions follow by standard arguments for a vector of sample means.

Next, note that, by Lemma A4 and iterated expectations,

$$\begin{aligned}
\Delta_\ell &= E\big[1(T_i(x^a)>0)g_{i1}(x^a)+B_\ell 1(T_i(x^a)=0)\big]\\
&\quad - E\big[1(T_i(x^b)>0)g_{i1}(x^b)+B_u 1(T_i(x^b)=0)\big]\\
&\le E\big[g_{i1}(x^a)\big]-E\big[g_{i1}(x^b)\big]=\Delta.
\end{aligned} \tag{S.4}$$

It follows similarly that $\Delta \le \Delta_u$. To show sharpness, let $\tilde{\alpha}_i = (\alpha_i, X_i)$. Define

$$\begin{aligned}
g(x,\tilde{\alpha}_i,\varepsilon_{it},C_a,C_b) &= 1(T_i(x)>0)g_0(x,\alpha_i,\varepsilon_{it})\\
&\quad + 1(T_i(x)=0)\big[C_a 1(x=x^a)+C_b 1(x=x^b)\big],
\end{aligned}$$

where $B_\ell \le C_a \le B_u$ and $B_\ell \le C_b \le B_u$. Note that $T_i(X_{it})>0$ with probability 1, so that $g(X_{it},\tilde{\alpha}_i,\varepsilon_{it},C_a,C_b) = g_0(X_{it},\alpha_i,\varepsilon_{it}) = Y_{it}$. Hence the conditional distribution of $(Y_{i1}, \ldots, Y_{iT})'$ given X_i is the same for g and $\tilde{\alpha}_i$ as for g_0 and α_i. Also, because (α_i, X_i) is a one-to-one function of $(\tilde{\alpha}_i, X_i)$, it follows that Assumption 2 is satisfied with $\tilde{\alpha}_i$ replacing α_i. When $(C_a, C_b) = (B_\ell, B_u)$, we have

$$\begin{aligned}\Delta &= E\big[g\big(x^a, \tilde{\alpha}_i, \varepsilon_{it}, B_\ell, B_u\big) - g\big(x^b, \tilde{\alpha}_i, \varepsilon_{it}, B_\ell, B_u\big)\big] \\ &= E\big[1\big(T_i\big(x^a\big) > 0\big)g_i\big(x^a\big) + 1\big(T_i\big(x^a\big) = 0\big)B_\ell\big] \\ &\quad - E\big[1\big(T_i\big(x^b\big) > 0\big)g_i\big(x^b\big) + 1\big(T_i\big(x^b\big) = 0\big)B_u\big] = \Delta_\ell,\end{aligned}$$

and the lower bound is attained. Similarly, the upper bound is attained when $(C_a, C_b) = (B_u, B_\ell)$.

Turning now to the quantile bounds, it follows as in the proof of Lemma A7 applied to $\hat{G}_\ell(y, x^a)$ and to $\hat{G}_\ell(y, x^b) + \overline{P}(x^b)$ that

$$\begin{aligned}&\hat{q}_u^d \xrightarrow{p} q_u^d, \quad \hat{q}_\ell^d \xrightarrow{p} q_\ell^d, \quad G_\ell\big(q_u^d, x^d\big) = \lambda, \\ &G_\ell\big(q_\ell^d, x^d\big) + \overline{\mathcal{P}}\big(x^d\big) = \lambda, \quad d \in \{a, b\}.\end{aligned}$$

It also follows as in equation (S.4) that $G_\ell(y, x) \le G(y, x) \le G_\ell(y, x) + \overline{\mathcal{P}}(x)$, implying $\Delta_{\lambda\ell} \le \Delta_\lambda \le \Delta_{\lambda u}$. Next, it follows as in Lemma A7 that

$$\begin{aligned}\sqrt{n}\big(\hat{q}_u^a - q_u^a\big) &= -G_\ell'\big(q_u^a, x^a\big)^{-1}\frac{1}{\sqrt{n}}\sum_i\big[G_i\big(q_u^a, x\big) - \lambda\big] + o_p(1), \\ \sqrt{n}\big(\hat{q}_\ell^b - q_\ell^b\big) &= -G_\ell'\big(q_\ell^b, x^b\big)^{-1} \\ &\qquad \times \frac{1}{\sqrt{n}}\sum_i\big[G_i\big(q_\ell^b, x^b\big) + 1\big(T_i\big(x^b\big) = 0\big) - \lambda\big] + o_p(1).\end{aligned}$$

Differencing then gives

$$\begin{aligned}&\sqrt{n}(\hat{\Delta}_u - \Delta_u) = -\sum_{i=1}^n \frac{\Psi_{\lambda i}^u}{\sqrt{n}} + o_p(1), \\ &\Psi_{\lambda i}^u = \frac{G_i(q_u^a, x^a) - \lambda}{G_\ell'(q_u^a, x^a)} - \frac{G_i(q_\ell^b, x^b) + 1(T_i(x^b) = 0) - \lambda}{G_\ell'(q_\ell^b, x^b)}.\end{aligned}$$

It follows similarly that

$$\begin{aligned}&\sqrt{n}(\hat{\Delta}_\ell - \Delta_\ell) = -\sum_{i=1}^n \frac{\Psi_{\lambda i}^\ell}{\sqrt{n}} + o_p(1), \\ &\Psi_{\lambda i}^\ell = \frac{G_i(q_\ell^a, x) + 1(T_i(x^a) = 0) - \lambda}{G_\ell'(q_\ell^a, x^a)} - \frac{G_i(q_u^b, x^b) - \lambda}{G_\ell'(q_u^b, x^b)}.\end{aligned}$$

Then, for $\Sigma_\lambda = \mathrm{Var}(\Psi^\ell_{\lambda i}, \Psi^u_{\lambda i})$, the next conclusion follows by the CLT. It also follows by arguments similar to the proof of Theorem A8 that $\sum_{i=1}^n (\hat{\Psi}^\ell_{\lambda i} - \Psi^\ell_{\lambda i})^2/n \xrightarrow{p} 0$ and $\sum_{i=1}^n (\hat{\Psi}^u_{\lambda i} - \Psi^u_{\lambda i})^2/n \xrightarrow{p} 0$. The consistency of $\hat{\Sigma}_\lambda$ then follows by standard methods.

To show sharpness of the QTE bounds, define $\tilde{\alpha}_i$ and $g(x, \tilde{\alpha}_i, \varepsilon_{it}, C_a, C_b)$ as in the proof of the ATE bounds, but now for any $C_a, C_b \in \mathbb{R}$. Let $G(y, x, C_a, C_b) = E[1(g(x, \tilde{\alpha}_i, \varepsilon_{it}, C_a, C_b) \le y)]$. Note that, for $d \in \{a, b\}$,

$$G(y, x^d, C_a, C_b) = G_\ell(y, x^d) + 1(y \ge C_d)\overline{\mathcal{P}}(x^d).$$

Let $q(\lambda, x, C_a, C_b)$ be the associated QSF. For $d \in \{a, b\}$,

$$\begin{aligned} &q(\lambda, x^d, C_a, C_b) \\ &\quad = \begin{cases} q_u(\lambda, x^d), & \lambda < G_\ell(C_d, x^d), \\ C_d, & G_\ell(C_d, x^d) \le \lambda \le G_\ell(C_d, x^d) + \overline{\mathcal{P}}(x^d), \\ q_\ell(\lambda, x^d), & \lambda > G_\ell(C_d, x^d) + \overline{\mathcal{P}}(x^d). \end{cases} \end{aligned}$$

For λ with $\overline{\mathcal{P}}(x^d) < \lambda < 1 - \overline{\mathcal{P}}(x^d)$, we have $q(\lambda, x^d, C_a, C_b) = q_\ell(\lambda, x^d)$ for C_d small enough that $G_\ell(C_d, x) + \overline{\mathcal{P}}(x^d) < \lambda$ and $q(\lambda, x^d, C_a, C_b) = q_u(\lambda, x^d)$ for C_d big enough. For $\lambda \le \overline{\mathcal{P}}(x^d)$, we have $q(\lambda, x^d, C_a, C_b) = q_u(\lambda, x)$ for all C_d big enough (by $\lambda < 1 - \overline{\mathcal{P}}(x^d)$) and $\lim_{C_d \longrightarrow -\infty} q(\lambda, x^d, C_a, C_b) = -\infty = q_\ell(\lambda, x)$. For $\lambda \ge 1 - \overline{\mathcal{P}}(x^d)$, we have $q(\lambda, x^d, C_a, C_b) = q_\ell(\lambda, x^d)$ for all C_d small enough and $\lim_{C_d \longrightarrow \infty} q(\lambda, x^d, C_a, C_b) = +\infty = q_u(\lambda, x^d)$. Therefore, we have

$$\begin{aligned} &\lim_{C_a \longrightarrow -\infty, C_b \longrightarrow +\infty} [q(\lambda, x^a, C_a, C_b) - q(\lambda, x^b, C_a, C_b)] \\ &\quad = q_\ell(\lambda, x^a) - q_u(\lambda, x^b), \\ &\lim_{C_a \longrightarrow +\infty, C_b \longrightarrow -\infty} [q(\lambda, x^a, C_a, C_b) - q(\lambda, x^b, C_a, C_b)] \\ &\quad = q_u(\lambda, x^a) - q_\ell(\lambda, x^b), \end{aligned}$$

showing the bounds are sharp. *Q.E.D.*

S.4.2. *Bounds Under Monotonicity*

We now turn to the bounds when g_0 is known to be monotonic, satisfying the following condition.

ASSUMPTION A1: *For some x^a and x^b, $g_0(x^a, \alpha_i, \varepsilon_{it}) \ge g_0(x^b, \alpha_i, \varepsilon_{it})$.*

This condition leads to tighter bounds for the ASF and QSF. Here we give results showing estimable population bounds under monotonicity. We also briefly describe how to estimate them, but for brevity do not give the full asymptotic theory. Define $1_i^a = 1(T_i(x^a) > 0)$, $1_i^b = 1(T_i(x^b) > 0)$, $\overline{\mathcal{P}}(x^b, x^a) = \Pr(T_i(x^a) = T_i(x^b) = 0)$, and

$$G_u^*(y, x^a) = E[G_i(y, x^a) + (1 - 1_i^a)G_i(y, x^b)] + \overline{\mathcal{P}}(x^b, x^a),$$
$$G_\ell^*(y, x^b) = E[G_i(y, x^b) + (1 - 1_i^b)G_i(y, x^a)].$$

THEOREM A9: *Suppose that Assumptions* 1, 2, 5, *and* A1 *are satisfied. If* $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ *for* $x \in \{x^a, x^b\}$ *then* $\Delta \geq P\delta$. *Also, if* $G_u^*(y, x^a)$ *and* $G_\ell^*(y, x^b)$ *are continuous and strictly increasing on the interior of their range, then* $q(\lambda, x^a) \geq Q(\lambda, G_u^*(\cdot, x^a))$ *and* $q(\lambda, x^b) \leq Q(\lambda, G_\ell^*(\cdot, x^b))$, *so that*

$$\Delta_\lambda \geq Q(\lambda, G_u^*(\cdot, x^a)) - Q(\lambda, G_\ell^*(\cdot, x^b)).$$

PROOF: Note that $1 = 1_i^a + (1 - 1_i^a)1_i^b + (1 - 1_i^a)(1 - 1_i^b)$. By Lemma A4,

$$E[1_i^a g_{i1}(x^a)] = E[\overline{Y}_i(x^a)], \quad E[1_i^b g_{i1}(x^b)] = E[\overline{Y}_i(x^b)].$$

Then by monotonicity,

$$\begin{aligned}
\mu(x^a) &= E[g_{i1}(x^a)] \\
&\geq E[\{1_i^a + (1 - 1_i^a)(1 - 1_i^b)\}g_{i1}(x^a)] + E[(1 - 1_i^a)1_i^b g_{i1}(x^b)] \\
&= E[1_i^a \overline{Y}_i(x^a) + (1 - 1_i^a)1_i^b \overline{Y}_i(x^b) + (1 - 1_i^a)(1 - 1_i^b)g_{i1}(x^a)].
\end{aligned}$$

Similarly,

$$\mu(x^b) \leq E[1_i^b \overline{Y}_i(x^b) + (1 - 1_i^b)1_i^a \overline{Y}_i(x^a) + (1 - 1_i^a)(1 - 1_i^b)g_{i1}(x^b)].$$

Subtracting this inequality from the previous one, and noting that $1_i^a - (1 - 1_i^b)1_i^a = 1_i^b 1_i^a = D_i$ and $-1_i^b + (1 - 1_i^a)1_i^b = -D_i$,

$$\begin{aligned}
&\mu(x^a) - \mu(x^b) \\
&\quad \geq E[D_i\{\overline{Y}_i(x^a) - \overline{Y}_i(x^b)\}] \\
&\quad\quad + E[(1 - 1_i^a)(1 - 1_i^b)\{g_0(x^a, \alpha_i, \varepsilon_{it}) - g_0(x^b, \alpha_i, \varepsilon_{it})\}] \\
&\quad \geq E[D_i\{\overline{Y}_i(x^a) - \overline{Y}_i(x^b)\}] = P\delta,
\end{aligned}$$

giving the first conclusion.

Next, similarly to above,

$$\begin{aligned}
G(y, x^a) &= E[\{1_i^a + (1 - 1_i^a)(1 - 1_i^b) + (1 - 1_i^a)1_i^b\}1(g_{i1}(x^a) \leq y)] \\
&\leq E[G_i(y, x^a)] + E[(1 - 1_i^a)G_i(y, x^b)] + \overline{\mathcal{P}}(x^b, x^a)
\end{aligned}$$

$$= G_u^*(y, x^a).$$

$$G(y, x^b) \geq G_\ell^*(y, x^b).$$

Inverting gives the second conclusion. *Q.E.D.*

Estimation of the bounds under monotonicity is straightforward. We can estimate the lower bound for the ATE by $(\sum_{i=1}^n D_i/n)\hat{\hat{\delta}}$. We can estimate the quantile bounds by inverting

$$\hat{G}_u^*(y, x^a) = \sum_{i=1}^n \big[\overline{G}_i(y, x^a) + (1 - 1_i^a)\overline{G}_i(y, x^b) + 1\big(T_i(x^b) = T_i(x^a) = 0\big)\big]/n,$$

$$\hat{G}_\ell^*(y, x^b) = \sum_{i=1}^n \big[\overline{G}_i(y, x^b) + (1 - 1_i^b)\overline{G}_i(y, x^a)\big]/n.$$

Asymptotic theory for these estimators of bounds under monotonicity is straightforward. We do not know if they are sharp.

S.5. SUPPLEMENTS TO SECTION 5

Here we give the proof of Theorem 3 as well as bounds that impose monotonicity.

S.5.1. *Proof of Theorem 3*

We first prove the second part of Lemma A4 for the dynamic model. Let $d_{it}(x) = 1(X_i \in \mathcal{X}_t(x))$. By Assumption 3, $\sum_{t=1}^T d_{it}(x) = 1(T_i(x) > 0)$, and the fact that $d_{it}(x)$ depends only on $X_{it}, X_{i,t-1}, \ldots, X_{i1}$, we have

$$\begin{aligned} E\big[\hat{Y}_i(x)|X_{i1}\big] &= \sum_{t=1}^T E\big[d_{it}(x)Y_{it}|X_{i1}\big] \\ &= \sum_{t=1}^T E\big[d_{it}(x)E\big[g_{it}(x)|X_{it}, \ldots, X_{i1}\big]|X_{i1}\big] \\ &= \sum_{t=1}^T E\big[d_{it}(x)E\big[g_{i1}(x)|X_{i1}\big]|X_{i1}\big] \\ &= E\big[1\big(T_i(x) > 0\big)|X_{i1}\big]E\big[g_{i1}(x)|X_{i1}\big]. \end{aligned}$$

Let

$$\begin{pmatrix} m_{\ell i} \\ m_{ui} \end{pmatrix} = \begin{pmatrix} \hat{Y}_i(x^a) - \hat{Y}_i(x^b) + B_\ell 1(T_i(x^a)=0) - B_u 1(T_i(x^b)=0) \\ \hat{Y}_i(x^a) - \hat{Y}_i(x^b) + B_u 1(T_i(x^a)=0) - B_\ell 1(T_i(x^b)=0) \end{pmatrix}.$$

Note that $\hat{\Delta}_\ell = \sum_{i=1}^n m_{\ell i}/n$ and $\hat{\Delta}_u = \sum_{i=1}^n m_{ui}/n$. Then, for $\Sigma = \text{Var}((m_{\ell i}, m_{ui}))$, $\Delta_\ell = E[m_{\ell i}]$, and $\Delta_u = E[m_{ui}]$, the first and second conclusions follow by standard arguments for a vector of sample means.

Next, note that $E[g_{i1}(x^a)|X_{i1}] \leq B_u$ by Assumption 6, so that

$$E[B_u 1(T_i(x^a)=0)|X_{i1}] \geq E[1(T_i(x^a)=0)|X_{i1}]E[g_{i1}(x^a)|X_{i1}].$$

Then by iterated expectations and $T_i(x^a) \geq 0$,

$$\begin{aligned} &E[\hat{Y}_i(x^a) + B_u 1(T_i(x^a)=0)|X_{i1}] \\ &\quad \geq E[1(T_i(x^a)>0)|X_{i1}]E[g_{i1}(x^a)|X_{i1}] \\ &\qquad + E[1(T_i(x^a)=0)|X_{i1}]E[g_{i1}(x^a)|X_{i1}] \\ &\quad = E[g_{i1}(x^a)|X_{i1}]. \end{aligned}$$

Taking expectations of both sides of this inequality gives

$$E[\hat{Y}_i(x^a) + B_u 1(T_i(x^a)=0)] \geq \mu(x^a).$$

Similarly, we have $E[\hat{Y}_i(x^a) + B_\ell 1(T_i(x^a)=0)] \leq \mu(x^a)$. Replacing x^a by x^b and differencing gives $\Delta_\ell \leq \Delta \leq \Delta_u$.

Turning to the quantile bounds, we next prove the first part of Lemma A4 for a dynamic model. Let $G_i(y,x)$ here, in the dynamic case, be given by

$$\begin{aligned} G_i(y,x) &= \sum_{t=1}^T d_{it}(x)1(Y_{it} \leq y) = \sum_{t=1}^T d_{it}(x)1(g_{it}(x) \leq y), \\ G_\ell(y,x) &= E[E[1(T_i(x)>0)|X_{i1}]1(g_{i1}(x) \leq y)]. \end{aligned}$$

Note that since $\sum_{t=1}^T d_{it}(x) = 1(T_i(x)>0)$ and $d_{it}(x)$ depends only on $X_{it}, X_{it-1}, \ldots, X_{i1}$, Assumption 3 implies

$$\begin{aligned} E[G_i(y,x)] &= E\left[\sum_{t=1}^T d_{it}(x)1(g_{it}(x) \leq y)\right] \\ &= E\left[\sum_{t=1}^T d_{it}(x)E[1(g_{it}(x) \leq y)|X_{it}, \ldots, X_{i1}]\right] \end{aligned}$$

$$
\begin{aligned}
&= E\left[\sum_{t=1}^{T} d_{it}(x) E\left[1\left(g_{i1}(x) \le y\right) | X_{i1}\right]\right] \\
&= E\left[1\left(T_i(x) > 0\right) E\left[1\left(g_{i1}(x) \le y\right) | X_{i1}\right]\right] \\
&= G_\ell(y, x).
\end{aligned}
$$

Also, since $d_{it}(x)d_{is}(x) = 0$ for any $s \neq t$ and $d_{it}(x)^2 = d_{it}(x)$, Assumption 3 implies that

$$
\begin{aligned}
&E\left[\left\{\hat{G}_i(y, x) - G_i(y, x)\right\}^2\right] \\
&\quad = E\left[\sum_{t=1}^{T} d_{it}(x)\left\{\Phi\left(\frac{y - g_{it}(x)}{h}\right) - 1\left(g_{it}(x) \le y\right)\right\}^2\right] \\
&\quad \le E\left[\sum_{t=1}^{T} d_{it}(x)\right. \\
&\qquad \left. \times E\left[\left\{\Phi\left(\frac{y - g_{it}(x)}{h}\right) - 1\left(g_{it}(x) \le y\right)\right\}^2 | X_{it}, \dots, X_{i1}\right]\right] \\
&\quad = E\left[1\left(T_i(x) > 0\right) E\left[\left\{\Phi\left(\frac{y - g_{i1}(x)}{h}\right) - 1\left(g_{i1}(x) \le y\right)\right\}^2 | X_{i1}\right]\right] \\
&\quad = E\left[E\left[1\left(T_i(x) > 0\right) | X_{i1}\right]\left\{\Phi\left(\frac{y - g_{i1}(x)}{h}\right) - 1\left(g_{i1}(x) \le y\right)\right\}^2\right].
\end{aligned}
$$

By Assumption 7 with X_{i1} replacing X_i, it follows that $g_{i1}(x)$ is continuously distributed for the probability measure weighted by $E[1(T_i(x) > 0)|X_{i1}]$. Therefore, it follows, similarly to the proof of Lemma A7, that $E[\{\hat{G}_i(y, x) - G_i(y, x)\}^2] \longrightarrow 0$ as $h \longrightarrow 0$. It also follows, similarly to the proof of Lemma A7, that

$$
E\left[\hat{G}_i(y, x)\right] = E\left[G_i(y, x)\right] + O\left(h^2\right).
$$

The conclusion now follows exactly like the proof of Theorem 2. *Q.E.D.*

S.5.2. *Bounds Under Monotonicity*

We now turn to the bounds when g_0 is known to be monotonic, satisfying Assumption A1, in the dynamic model. This condition leads to tighter bounds for the ASF and QSF. Here we give results showing estimable population bounds under monotonicity. We also briefly describe how to estimate them, but for brevity do not give the full asymptotic theory. For $d \in \{a, b\}$, define

$1^d_{it} = 1(X_i \in \mathcal{X}_t(x^d))$, $t = 1, \ldots, T$, $\overline{1}^d_i = 1(X_i \in \overline{\mathcal{X}}(x^d))$, and $\tilde{1}^d_{iT} = 1(X_{iT} = x^d)$. Let

$$G^*_u(y, x^a) = E\big[G_i(y, x^a) + \overline{1}^a_i\big\{\tilde{1}^b_{iT}1(Y_{iT} \le y) + \big(1 - \tilde{1}^b_{iT}\big)\big\}\big],$$
$$G^*_\ell(y, x^b) = E\big[G_i(y, x^b) + \overline{1}^b_i\tilde{1}^a_{iT}1(Y_{iT} \le y)\big].$$

THEOREM A10: *Suppose that Assumptions* 1, 3, 5, *and* A1 *are satisfied*. *If* $E[|g_0(x, \alpha_i, \varepsilon_{it})|] < \infty$ *for* $x \in \{x^a, x^b\}$, *then*

$$\begin{aligned}\Delta \ge{}& E\big[\hat{Y}_i(x^a) - \hat{Y}_i(x^b)\big] + E\big[\overline{1}^a_i\big(\tilde{1}^b_{iT}Y_{iT} + \big(1 - \tilde{1}^b_{iT}\big)B_\ell\big)\big] \\ &- E\big[\overline{1}^b_i\big(\tilde{1}^a_{iT}Y_{iT} + \big(1 - \tilde{1}^a_{iT}\big)B_u\big)\big].\end{aligned}$$

Also, *if* $G^*_u(y, x^a)$ *and* $G^*_\ell(y, x^b)$ *are continuous and strictly increasing on the interior of their range*, *then* $q(\lambda, x^a) \ge Q(\lambda, G^*_u(\cdot, x^a))$ *and* $q(\lambda, x^b) \le Q(\lambda, G^*_\ell(\cdot, x^b))$, *so that*

$$\Delta_\lambda \ge Q\big(\lambda, G^*_u\big(\cdot, x^a\big)\big) - Q\big(\lambda, G^*_\ell\big(\cdot, x^b\big)\big).$$

PROOF: Note that $1 = \sum_{t=1}^T 1^a_{it} + \overline{1}^a_i\tilde{1}^b_{iT} + \overline{1}^a_i(1 - \tilde{1}^b_{iT})$. By Lemma A4,

$$\sum_{t=1}^T E\big[1^a_{it}g_{it}\big(x^a\big)\big] = E\big[\hat{Y}_i\big(x^a\big)\big], \quad \sum_{t=1}^T E\big[1^b_{it}g_{it}\big(x^b\big)\big] = E\big[\hat{Y}_i\big(x^b\big)\big].$$

Then by Assumption 3, monotonicity, and $g_{iT}(x^a) \ge B_\ell$,

$$\begin{aligned}\mu\big(x^a\big) &= \sum_{t=1}^T E\big[1^a_{it}g_{it}\big(x^a\big)\big] + E\big[\overline{1}^a_i g_{iT}\big(x^a\big)\big] \\ &\ge \sum_{t=1}^T E\big[1^a_{it}g_{it}\big(x^a\big)\big] + E\big[\overline{1}^a_i\tilde{1}^b_{iT}g_{iT}\big(x^b\big)\big] + E\big[\overline{1}^a_i\big(1 - \tilde{1}^b_{iT}\big)\big]B_\ell \\ &= E\big[\hat{Y}_i\big(x^a\big)\big] + E\big[\overline{1}^a_i\tilde{1}^b_{iT}Y_{iT}\big] + E\big[\overline{1}^a_i\big(1 - \tilde{1}^b_{iT}\big)\big]B_\ell.\end{aligned}$$

Similarly, we have

$$\mu\big(x^b\big) \le E\big[\hat{Y}_i\big(x^b\big)\big] + E\big[\overline{1}^b_i\tilde{1}^a_{iT}Y_{iT}\big] + E\big[\overline{1}^b_i\big(1 - \tilde{1}^a_{iT}\big)\big]B_u.$$

Subtracting this inequality from the previous one gives the first conclusion.

Next, similarly to above,

$$\begin{aligned} G(y, x^a) &= \sum_{t=1}^{T} E\big[1_{it}^a 1\big(g_{it}(x^a) \le y\big)\big] + E\big[\overline{1}_i^a 1\big(g_{iT}(x^a) \le y\big)\big] \\ &\le E\big[G_i(y, x^a)\big] + E\big[\overline{1}_i^a \tilde{1}_{iT}^b 1(Y_{iT} \le y)\big] + E\big[\overline{1}_i^a\big(1 - \tilde{1}_{iT}^b\big)\big] \\ &= G_u^*(y, x^a), \\ G(y, x^b) &\ge G_\ell^*(y, x^b). \end{aligned}$$

Inverting gives the second conclusion. *Q.E.D.*

If $X_{it} \in \{0, 1\}$, $x^b = 0$, and $x^a = 1$, then $\overline{1}_i^b(1 - \tilde{1}_{iT}^a) = \overline{1}_i^a(1 - \tilde{1}_{iT}^b) = 0$ and the lower bound for Δ does not depend on B_ℓ and B_u.

Estimation of the bounds under monotonicity is straightforward. We can estimate the lower bound for the ATE by

$$\sum_{i=1}^{n} \big[\hat{Y}_i(x^a) - \hat{Y}_i(x^b) + \overline{1}_i^a\big(\tilde{1}_{iT}^b Y_{iT} + \big(1 - \tilde{1}_{iT}^b\big)B_\ell\big) \\ - \overline{1}_i^b\big(\tilde{1}_{iT}^a Y_{iT} + \big(1 - \tilde{1}_{iT}^a\big)B_u\big)\big]/n.$$

We can estimate the quantile bounds by inverting

$$\hat{G}_u^*(y, x^a) = \sum_{i=1}^{n} \big[\hat{G}_i(y, x^a) + \overline{1}_i^a\big\{\tilde{1}_{iT}^b 1(Y_{iT} \le y) + \big(1 - \tilde{1}_{iT}^a\big)\big\}\big]/n,$$

$$\hat{G}_\ell^*(y, x^b) = \sum_{i=1}^{n} \big[\hat{G}_i(y, x^b) + \overline{1}_i^b \tilde{1}_{iT}^a 1(Y_{iT} \le y)\big]/n.$$

Asymptotic theory for these estimators of bounds under monotonicity is straightforward. We do not know if they are sharp.

S.6. SUPPLEMENTS TO SECTION 6

In addition to the proofs of the rate results of Section 6, we here give necessary and sufficient conditions for identification as $T \longrightarrow \infty$ and extend the identification and rate results to the QTE.

S.6.1. *Identification as $T \to \infty$*

We begin with the identification result. The necessary and sufficient condition for identification of Δ as T grows is the following.

ASSUMPTION A2: $\Pr(\Pr(X_{it} = x|\alpha_i) > 0) = 1$ *for* $x \in \{x^a, x^b\}$ *and some* $t \in \{1, \dots, T\}$.

If this condition does not hold for both x^b and x^a, then some individuals, as represented by α_i, will never reach either x^b or x^a, so we cannot nonparametrically identify the treatment effect for those individuals, and hence the overall treatment effect is not identified. A related condition was formulated in Chamberlain (1982, p. 17) but was used for a different purpose, as a sufficient condition for a least-squares estimate for a single individual to converge to that individual's coefficient.

The following result shows the key role of Assumption A2 in achieving identification as $T \longrightarrow \infty$.

THEOREM A11: *Suppose that Assumptions* 1 *and* 5 *are satisfied. If Assumption* A2 *is not satisfied, then* $\overline{\mathcal{P}}(x)$ *is bounded away from zero uniformly in* T *for* $x = x^a$ *or* $x = x^b$, *so that if Assumption* 6 *is satisfied,* $\Delta_u - \Delta_\ell$ *does not converge to zero as* T *grows. Suppose also that* $(X_{i1}, X_{i2}, \dots)$ *is stationary and ergodic conditional on* α_i. *If Assumptions* 2 *and* A2 *are satisfied and* $E[|g_0(x, \alpha_i, \varepsilon_{i1})|] < \infty$ *for* $x = x^a$ *and* $x = x^b$, *then* $\delta \longrightarrow \Delta$ *as* $T \longrightarrow \infty$. *If Assumptions* 3, 6, *and* A2 *are satisfied, then* $\Delta_u - \Delta_\ell \longrightarrow 0$ *as* $T \longrightarrow \infty$.

PROOF: First, note that if Assumption A2 is not satisfied, then, for some $x^d \in \{x^a, x^b\}$, there is a set $\mathcal{A}$ with $\Pr(\mathcal{A}) > 0$ such that $\Pr(X_{it} = x^d|\alpha_i) = 0$ for all t and $\alpha_i \in \mathcal{A}$. Then

$$E\big[T_i\big(x^d\big)|\alpha_i \in \mathcal{A}\big] = \sum_{t=1}^{T} E\big[1\big(X_{it} = x^d\big)|\alpha_i \in \mathcal{A}\big] = 0.$$

Since $T_i(x^d)$ is a nonnegative random variable, this implies that $\Pr(T_i(x^d) = 0|\alpha_i) = 1$ for all T and $\alpha_i \in \mathcal{A}$. Therefore

$$\begin{aligned}\overline{\mathcal{P}}\big(x^d\big) &= E\big[\Pr\big(T_i\big(x^d\big) = 0|\alpha_i\big)\big] \geq E\big[1(\mathcal{A}) \Pr\big(T_i\big(x^d\big) = 0|\alpha_i\big)\big] \\ &= \Pr(\mathcal{A}) > 0.\end{aligned}$$

Thus $\overline{\mathcal{P}}(x^d)$ is bounded away from zero for all T, and hence, under Assumption 6, $(B_u - B_\ell)[\overline{\mathcal{P}}(x^a) + \overline{\mathcal{P}}(x^b)] \geq (B_u - B_\ell)\overline{\mathcal{P}}(x^d)$ does not converge to zero.

Next suppose that Assumptions 2 and A2 are satisfied, $(X_{i1}, X_{i2}, \dots)$ is stationary and ergodic conditional on α_i, and that $x \in \{x^a, x^b\}$. Recall that $T_i(x) = \sum_{t=1}^{T} 1(X_{it} = x)$. By the ergodic theorem, there is a set of α_i having probability 1 such that

$$T_i(x)/T \xrightarrow{\text{a.s.}} E\big[1(X_{it} = x)|\alpha_i\big] = \Pr(X_{it} = x|\alpha_i).$$

Under Assumption A2, $\Pr(X_{it} = x|\alpha_i) > 0$ on a set of α_i with probability 1 (a.s. α_i henceforth). Therefore $1(T_i(x) > 0) \xrightarrow{\text{a.s.}} 1$ a.s. α_i. Since this holds for both x^a and x^b, it follows that

$$D_i = 1\big(T_i\big(x^a\big) > 0\big)1\big(T_i\big(x^b\big) > 0\big) \xrightarrow{\text{a.s.}} 1$$

a.s. α_i. Let $\Delta_i = g_{i1}(x^a) - g_{i1}(x^b)$. Note that $|D_i\Delta_i| \leq |\Delta_i|$ and $E[|\Delta_i||\alpha_i] < \infty$ a.s. α_i. Then by the dominated convergence theorem (DCT henceforth),

$$E[D_i\Delta_i|\alpha_i] \longrightarrow E[\Delta_i|\alpha_i], \quad E[D_i|\alpha_i] \longrightarrow 1 \quad \text{a.s.} \quad \alpha_i.$$

Then, by applying the DCT again,

$$E[D_i\Delta_i] \longrightarrow E[\Delta_i] = \Delta, \quad E[D_i] \longrightarrow 1,$$

giving the first conclusion.

Suppose next that Assumptions 3 and 6 are satisfied, and $(X_{i1}, X_{i2}, \ldots)$ is stationary and ergodic conditional on α_i. Recall that $\Delta_u - \Delta_\ell = (B_u - B_\ell)[\overline{\mathcal{P}}(x^a) + \overline{\mathcal{P}}(x^b)]$. If Assumption A2 is satisfied, then, since $1(T_i(x^a) > 0) \geq D_i$, we have

$$\overline{\mathcal{P}}\big(x^a\big) = E\big[1\big(T_i\big(x^a\big) = 0\big)\big] \leq 1 - E[D_i] \longrightarrow 0.$$

Similarly, we have $\overline{\mathcal{P}}(x^b) \longrightarrow 0$ so the second conclusion holds. *Q.E.D.*

S.6.2. *Proof of Theorem 4*

Let $\prod_{t=1}^T 1(X_{it} \neq x)$ be the indicator function for the event that none of the elements of X_i is equal to x, so that $\overline{\mathcal{P}}(x) = E[\prod_{t=1}^T 1(X_{it} \neq x)]$. By iterated expectations, for $T > J$,

$$\begin{aligned}
\overline{\mathcal{P}}(x) &= E\left[\prod_{t=1}^{T-1} 1(X_{it} \neq x) E\big[1(X_{iT} \neq x)|X_{i,T-1}, \ldots, X_{i1}, \alpha_i\big]\right] \\
&= E\left[\left\{\prod_{t=1}^{T-1} 1(X_{it} \neq x)\right\} \Pr(X_{iT} \neq x|X_{i,T-1}, \ldots, X_{i,T-J}, \alpha_i)\right] \\
&\leq (1-\varepsilon) E\left[\prod_{t=1}^{T-1} 1(X_{it} \neq x)\right].
\end{aligned}$$

Repeating the argument for $T-1, \ldots, J$ gives

$$\overline{\mathcal{P}}(x) \leq (1-\varepsilon)^{T-J} E\left[\prod_{t=1}^{J-1} 1(X_{it} \neq x)\right] \leq (1-\varepsilon)^{T-J},$$

giving the first conclusion.

For the second conclusion, note that the conditional i.i.d. assumption and the bound imply that, for $P_i = \Pr(X_{it} \neq x|\alpha_i)$, we have $\overline{\mathcal{P}}(x) = E[P_i^T]$ being no greater than a constant times the Tth raw moment of a Beta distribution with parameters γ and v. Also, it is well known that $T^v\Gamma(T+\gamma)/\Gamma(T+\gamma+v) \longrightarrow 1$ as $T \longrightarrow \infty$. Therefore, we have

$$\begin{aligned} E\big[P_i^T\big] &\le C\big[\Gamma(\gamma+v)/\Gamma(\gamma)\Gamma(v)\big]\int_0^1 p^{T+\gamma-1}(1-p)^{v-1}\,dp \\ &\le C\big[\Gamma(\gamma+v)/\Gamma(\gamma)\Gamma(v)\big]\big[\Gamma(T+\gamma)\Gamma(v)/\Gamma(T+\gamma+v)\big] \\ &= C\Gamma(T+\gamma)/\Gamma(T+\gamma+v) \le CT^{-v}. \qquad \textit{Q.E.D.} \end{aligned}$$

S.6.3. *Proof of Theorem 5*

Note that $\Pr(Y_{it}=0|Y_{i,t-1}=0,\alpha_i) = 1 - H(\alpha_{1i})$,

$$\begin{aligned} \overline{\mathcal{P}}(1) &= E\big[\Pr(Y_{i,T-1} = Y_{i,T-2} = \cdots = Y_{i0} = 0|\alpha_i)\big] \\ &= E\left[\prod_{t=1}^{T-1}\Pr(Y_{it}=0|Y_{i,t-1}=0,\alpha_i)\Pr(Y_{i0}=0|\alpha_i)\right] \\ &\le E\big[\big\{1-H(\alpha_{i1})\big\}^{T-1}\big]. \end{aligned}$$

By a change of variables, we find that the p.d.f. $f(p)$ of $1-H(\alpha_{i1})$ is

$$f(p) = f_1\big(H^{-1}(1-p)\big)/f_\varepsilon\big(H^{-1}(1-p)\big) \le C(1-p)^{v-1}p^{v-1}.$$

Thus, the p.d.f. of $1-H(\alpha_{i1})$ is bounded above by a Beta p.d.f. with parameters v, v. It then follows, as in the proof of Theorem 4, that $\overline{\mathcal{P}}(1) \le C(T-1)^{-v} \le CT^{-v}$. It follows similarly that $\overline{\mathcal{P}}(0) \le CT^{-v}$. *Q.E.D.*

S.6.4. *Identification Rates for QTE*

Finally, we show that the nonparametric rates and nonidentification results apply to the QTE. We do this by giving lemmas for quantile bounds that apply to both static and dynamic models. The first lemma shows that the identification rate is at least as fast as the rate at which $\overline{\mathcal{P}}(x)$ decreases.

LEMMA A12: *Suppose that $G(y)$ is a CDF that is strictly increasing and continuously differentiable on $\{y : 0 < G(y) < 1\}$ and that $G_T(y)$ is a continuous function and $\overline{\mathcal{P}}_T$ a nonnegative constant satisfying*

$$G_T(y) \le G(y) \le G_T(y) + \overline{\mathcal{P}}_T, \quad G_T(-\infty) = 0, \quad G_T(\infty) + \overline{\mathcal{P}}_T = 1.$$

If $\overline{\mathcal{P}}_T \longrightarrow 0$ *as* $T \longrightarrow \infty$, *then*, *for* $0 < \lambda < 1$ *and large enough* T, *there are* $q_{\ell T} \le q \le q_{uT}$ *satisfying*

$$\lambda = G_T(q_{uT}) = G(q) = G_T(q_{\ell T}) + \overline{\mathcal{P}}_T.$$

Also, *any such* q_{uT} *and* $q_{\ell T}$ *satisfy* $q_{uT} - q_{\ell T} = O(\overline{\mathcal{P}}_T)$.

PROOF: Choose T large enough that $\overline{\mathcal{P}}_T < \min(\lambda, 1-\lambda)$. Then $G_T(\infty) = 1 - \overline{\mathcal{P}}_T > \lambda$ and $G_T(-\infty) + \overline{\mathcal{P}}_T = \overline{\mathcal{P}}_T < \lambda$. Therefore, by continuity of $G_T(y)$, there exist q_{uT} such that $\lambda = G_T(q_{uT})$ and $q_{\ell T}$ such that $\lambda = G_T(q_{\ell T}) + \overline{\mathcal{P}}_T$. Also, by $G(y)$ being a strictly increasing CDF, there is a unique q with $\lambda = G(q)$. Note that $G(q) = G_T(q_{uT}) \le G(q_{uT})$, so that $q_{uT} \ge q$ by $G(q)$ strictly monotonic. It follow similarly that $q_{\ell T} \le q$. Also, for any $\varepsilon > 0$, we have $G(q - \varepsilon) < G(q)$, so that, for large enough T, it follows that

$$G(q-\varepsilon) < G(q) - \overline{\mathcal{P}}_T = G_T(q_{\ell T}) \le G(q_{\ell T}).$$

By strict monotonicity of $G(q)$, it follows that $q_{\ell T} > q - \varepsilon$ for large enough T. Since ε is arbitrary, we have $q_{\ell T} \longrightarrow q$. It follow similarly that $q_{uT} \longrightarrow q$.

Next, choose ε small enough that $\partial G(\tilde{q})/\partial q \ge C > 0$ for $\tilde{q} \in \mathcal{I} = [q-\varepsilon, q+\varepsilon]$. Note that, for T large enough, $q_{\ell T}, q_{uT} \in \mathcal{I}$. Also we have

$$\begin{aligned} G(q_{\ell T}) + 2\overline{\mathcal{P}}_T &\ge G_T(q_{\ell T}) + 2\overline{\mathcal{P}}_T = G(q) + \overline{\mathcal{P}}_T \\ &= G_T(q_{uT}) + \overline{\mathcal{P}}_T \ge G(q_{uT}). \end{aligned}$$

Subtracting $G(q_{\ell T})$ from both sides and expanding gives

$$2\overline{\mathcal{P}}_T \ge G(q_{uT}) - G(q_{\ell T}) = \frac{\partial G(\overline{q}_T)}{\partial q}(q_{uT} - q_{\ell T}) \ge C(q_{uT} - q_{\ell T}).$$

Dividing through by C gives $q_{uT} - q_{\ell T} \le C\overline{\mathcal{P}}_T$, implying the conclusion. *Q.E.D.*

The next result gives conditions under which the identification rate is no faster than the rate at which $\overline{\mathcal{P}}(x)$ decreases. This result will also show that quantile effects are not identified as $T \longrightarrow \infty$ if $\overline{\mathcal{P}}(x)$ does not go to zero.

LEMMA A13: *If the conditions of Lemma A12 are satisfied and* $G_T(y)$ *is continuously differentiable with* $|\partial G_T(y)/\partial y| \le C$ *for all* y *and* T, *then there is* C *such that*, *for* $\overline{\mathcal{P}}_T > 0$,

$$q_{uT} - q_{\ell T} \ge C\overline{\mathcal{P}}_T.$$

PROOF: As in the proof of Lemma A12, we have $G_T(q_{uT}) = G_T(q_{\ell T}) + \overline{\mathcal{P}}_T$. By the intermediate value theorem, it follows that, for some $q_{\ell T} \leq \overline{q} \leq q_{uT}$,

$$\frac{\partial G_T(\overline{q})}{\partial q}(q_{uT} - q_{\ell T}) = \overline{\mathcal{P}}_T.$$

For $\overline{\mathcal{P}}_T > 0$, we must have $\partial G_T(\overline{q})/\partial q \neq 0$, so that

$$q_{uT} - q_{\ell T} = \left[\frac{\partial G_T(\overline{q})}{\partial q}\right]^{-1} \overline{\mathcal{P}}_T \geq C^{-1}\overline{\mathcal{P}}_T. \qquad \textit{Q.E.D.}$$

Taken together, these two results show that the identification rate for the QTE is the same as the rate at which $\overline{\mathcal{P}}(x)$ decreases. Together they also show that if $\overline{\mathcal{P}}(x)$ does not go to zero, the bounds do not shrink to a point. It is straightforward to check that the conditions of these results are satisfied.

S.7. SUPPLEMENTS TO SECTION 7

We now turn to the results of Section 7 and to one additional result on the consistency of non-linear fixed-effects estimators for the identified ATE.

S.7.1. *Proof of Theorem 6*

Consider first the static case where $X_{it} \in \{0, 1\}$. We show the result for $X^k = (0, \dots, 0)'$. The result for $X^k = (1, \dots, 1)'$ will follow similarly. Note that β^* is identified for logit so $B = \{\beta^*\}$. Let $Z = H(\alpha)$ and let $G(z)$ be the CDF of Z when $F \in \mathcal{F}_k = \mathcal{F}_k(\beta^*, \mathcal{P})$ is the CDF of α. By $(Y_{i1}, \dots, Y_{iT})$ mutually independent conditional on α, we have

$$\begin{aligned} M_t &= \Pr\bigl(Y_{it} = 1, \dots, Y_{i1} = 1 | X_i \in X^k\bigr) \\ &= \int H(\alpha)^t \, dF(\alpha) = \int Z^t \, dG(Z), \end{aligned}$$

so that M_t is identified for $t = 1, \dots, T$. Now consider a Tth-order polynomial $P(z, T) = b_0 + b_1 z + \cdots + b_T z^T$ in z. Note that

$$\int P(Z, T)\, dG(Z) = b_0 + \sum_{t=1}^{T} b_t M_t$$

does not depend on $F \in \mathcal{F}_k$. As a special case, $\int Z\, dG(Z) = M_1$ also does not depend on $F \in \mathcal{F}_k$. Define the function $h(z) = H(\beta^* + H^{-1}(z)) = ze^{\beta^*}/1 - (1 - e^{\beta^*})z$. Note that $\Delta^k = \int [h(Z) - Z]\, dG(Z)$ for all $F \in \mathcal{F}_k$. For any polynomial

$P(z,t)$, let $R(z,t)=h(z)-P(z,t)$ be the remainder. Then we have

$$
\begin{aligned}
\text{(S.5)}\qquad & \Delta_u^k-\Delta_\ell^k \\
&= \sup_{F\in\mathcal{F}_k}\int\big[h(Z)-Z\big]\,dG(Z)-\inf_{F\in\mathcal{F}_k}\int\big[h(Z)-Z\big]\,dG(Z) \\
&= \sup_{F\in\mathcal{F}_k}\int\big[P(Z,T)+R(Z,T)\big]\,dG(Z) \\
&\quad -\inf_{F\in\mathcal{F}_k}\int\big[P(Z,T)+R(Z,T)\big]\,dG(Z) \\
&= \sup_{F\in\mathcal{F}_k}\int R(Z,T)\,dG(Z)-\inf_{F\in\mathcal{F}_k}\int R(Z,T)\,dG(Z) \\
&\le 2\sup_{0\le z\le 1}\big|R(z,T)\big|.
\end{aligned}
$$

The function $h(z)$ is continuously differentiable of order r for every r with

$$
\left|\frac{d^r h(z)}{dz^r}\right|\le r!e^{|\beta^*|}\big(e^{|\beta^*|}-1|\big)^{r-1}.
$$

Then, by Jackson's theorem (e.g., Judd (1998), chap. 3), there exists $P(z,T)$ such that, for $\gamma=\pi(e^{|\beta^*|}-1|)/4$,

$$
\begin{aligned}
\sup_{0\le z\le 1}\big|R(z,T)\big| &\le \frac{(T-r)!}{T!}\left(\frac{\pi}{4}\right)^r\sup_{0\le z\le 1}\left|\frac{d^r h(z)}{dz^r}\right| \\
&\le \frac{(T-r)!r!}{T!}\left(\frac{\pi}{4}\right)^r e^{|\beta^*|}\big(e^{|\beta^*|}-1|\big)^{r-1}\le C\left(\frac{r\gamma}{T}\right)^r.
\end{aligned}
$$

This inequality continues to hold if γ is replaced by $\max\{\gamma,1\}$, so we can assume $\gamma>1$. Then choose r equal to $T/\gamma e$, so that

$$
\sup_{0\le z\le 1}\big|R(z,T)\big|\le Ce^{-T/\gamma e}.
$$

The conclusion then follows by equation (S.5).

Next consider the dynamic binary-logit model where $X_{it}=Y_{i,t-1}$. It is known from Cox (1958) and Chamberlain (1985) that β^* is identified for T large enough. We show the result for Δ^1 where $\mathcal{X}^1=\{X_i:X_{i1}=0\}$. The result for the ATE conditional on $X_{i1}=1$ will follow analogously. Here

$$
\Pr(Y_{it}=0,\dots,Y_{i1}=0|X_{i1}=0)=\int\big[1-H(\alpha)\big]^t\,dF(\alpha)
$$

is identified for $t = 1, \dots, T$. It follows by a standard argument that $M_t = \int H(\alpha)^t \, dF(\alpha)$ is identified for $t = 1, \dots, T$. The proof then proceeds exactly as for the static case. *Q.E.D.*

S.7.2. *Consistency of Fixed Effects for Identified ATE*

We now consider the fixed-effects estimator in a binary-choice model with a binary regressor and $T = 2$. In some models, fixed-effects (FE) estimators of the ATE appear to have small biases; for example, see Hahn and Newey (2004) and Fernández-Val (2009). Here we show consistency of FE for δ. To describe this result, note that the FE estimator of the ASF conditional on $X_i = X^k$ is

$$\hat{\mu}_k^{\mathrm{FE}}(x) = \sum_{i=1}^{n} 1\big(X_i = X^k\big) H(x\hat{\beta}_{\mathrm{FE}} + \hat{\alpha}_i) \Big/ \sum_{i=1}^{n} 1\big(X_i = X^k\big),$$

$$\hat{\beta}_{\mathrm{FE}}, \hat{\alpha}_1, \dots, \hat{\alpha}_n = \arg\max_{\beta, \alpha_1, \dots, \alpha_n} \sum_{i,t} \ln\big\{H(X_{it}\beta + \alpha_i)^{Y_{it}} \times \big[1 - H(X_{it}\beta + \alpha_i)\big]^{1-Y_{it}}\big\}.$$

Let β_T denote the limit of $\hat{\beta}_{\mathrm{FE}}$. In the multinomial choice model, $\hat{\alpha}_i$ have a limit distribution conditional on $X_i = X^k$ that is discrete with J support points $\alpha_j^k(\beta_T)$ and $\Pr(\alpha = \alpha_j^k(\beta_T)) = \mathcal{P}_j^k$ $(j = 1, \dots, J)$. These limits satisfy

$$\text{(S.6)} \qquad \beta_T = \arg\max_{\beta} \sum_{k=1}^{K} \mathcal{P}^k \sum_{j=1}^{J} \mathcal{P}_j^k \log \mathcal{L}_j^k\big(\alpha_j^k(\beta), \beta\big),$$

$$\alpha_j^k(\beta) = \arg\max_{\alpha} \mathcal{L}_j^k(\alpha, \beta) \qquad (j = 1, \dots, J; k = 1, \dots, K),$$

where $\mathcal{P}^k = E[1(X_i = X^k)]$. The corresponding limit of $\hat{\mu}_k^{\mathrm{FE}}(x)$ is then given by

$$\mu_k^T(x) = \sum_{j=1}^{J} \mathcal{P}_j^k H\big(x'\beta_T + \alpha_j^k(\beta_T)\big).$$

Note that, with binary X_{it} and $T = 2$, we have $K = 4$. Let $X^1 = (0, 0)$, $X^2 = (0, 1)$, $X^3 = (1, 0)$, and $X^4 = (1, 1)$, so that the identified effect equals $\delta = \sum_{k=2}^{3} \mathcal{P}^k \Delta^k / \sum_{k=2}^{3} \mathcal{P}^k$.

THEOREM A14: *If* $H'(x) > 0$, $H(-x) = 1 - H(x)$, $X_{it} \in \{0, 1\}$, $T = 2$, *and* $\mathcal{P}_2 + \mathcal{P}_3 > 0$, *then*

$$\sum_{k=2}^{3} \mathcal{P}^k \big[\mu_k^T(1) - \mu_k^T(0)\big] \Big/ \sum_{k=2}^{3} \mathcal{P}^k = \delta.$$

PROOF: Let $Y^1 = (0,0)'$, $Y^2 = (0,1)'$, $Y^3 = (1,0)'$, $Y^4 = (1,1)'$, and $X^1 = (0,0)'$, $X^2 = (0,1)'$, $X^3 = (1,0)'$, $X^4 = (1,1)'$. The identified effect is

$$\begin{aligned}\delta &= \{\mathcal{P}^2 E[Y_{i2} - Y_{i1} | X_i = X^2] + \mathcal{P}^3 E[Y_{i1} - Y_{i2} | X_i = X^3]\} / (\mathcal{P}^2 + \mathcal{P}^3) \\ &= [\mathcal{P}^2(\mathcal{P}_2^2 - \mathcal{P}_3^2) + \mathcal{P}^3(\mathcal{P}_3^3 - \mathcal{P}_2^3)] / (\mathcal{P}^2 + \mathcal{P}^3).\end{aligned}$$

Next, the symmetry $H(-x) = 1 - H(x)$ implies that $\alpha_j^k(\beta)$ take the form

$$\alpha_j^k(\beta) = \begin{cases} -\infty, & j = 1, \\ -\beta(X_1^k + X_2^k)/2, & j = 2, 3, \\ \infty, & j = 4. \end{cases}$$

Note that, for $k = 2$ or $k = 3$, we have $X_1^k + X_2^k = 1$, so that $\alpha_j^k(\beta) = -\tilde{\beta}$ for $\tilde{\beta} = \beta/2$. Thus,

$$H(\beta + \alpha_j^k(\beta)) - H(\alpha_j^k(\beta)) = H(\tilde{\beta}) - H(-\tilde{\beta}) = 2H(\tilde{\beta}) - 1.$$

Therefore, the limit of the fixed-effects estimator of the identified effect is

$$A[2H(\tilde{\beta}) - 1], \quad A = [\mathcal{P}^2(\mathcal{P}_2^2 + \mathcal{P}_3^2) + \mathcal{P}^3(\mathcal{P}_2^3 + \mathcal{P}_3^3)] / (\mathcal{P}^2 + \mathcal{P}^3).$$

Next, the limit of the concentrated log-likelihood is

$$2\mathcal{P}^2[\mathcal{P}_2^2 \ln H(\tilde{\beta}) + \mathcal{P}_3^2 \ln H(-\tilde{\beta})] + 2\mathcal{P}^3[\mathcal{P}_2^3 \ln H(-\tilde{\beta}) + \mathcal{P}_3^3 \ln H(\tilde{\beta})].$$

The first-order conditions for maximization of this object are

$$0 = 2\mathcal{P}^2[\mathcal{P}_2^2 \lambda(\tilde{\beta}) - \mathcal{P}_3^2 \lambda(-\tilde{\beta})] + 2\mathcal{P}^3[-\mathcal{P}_2^3 \lambda(-\tilde{\beta}) + \mathcal{P}_3^3 \lambda(\tilde{\beta})],$$

where $\lambda(x) = H'(x)/H(x)$. By symmetry, $H'(-\tilde{\beta}) = H'(\tilde{\beta})$. Divide the first-order conditions by $H'(\tilde{\beta})$ and multiply by $H(\tilde{\beta})H(-\tilde{\beta})$ to obtain

$$\begin{aligned}0 &= 2\mathcal{P}^2[\mathcal{P}_2^2 H(-\tilde{\beta}) - \mathcal{P}_3^2 H(\tilde{\beta})] + 2\mathcal{P}^3[-\mathcal{P}_2^3 H(\tilde{\beta}) + \mathcal{P}_3^3 H(-\tilde{\beta})] \\ &= 2(\mathcal{P}^2 + \mathcal{P}^3)[\delta - A(2H(\tilde{\beta}) - 1)].\end{aligned}$$

Q.E.D.

In numerical examples, this same result continues to hold for $T = 3$ and $T = 4$. It would be interesting to extend this result to larger T, but it is beyond the scope of this paper to do so. Unfortunately, this result does not extend to the overall ATE.

S.8. SUPPLEMENTS TO SECTION 8

Here we give the proofs of Section 8 and additional numerical results for the logit model.

S.8.1. *Proof of Lemma 7*

Let the vector of model probabilities for $(Y^1, \dots, Y^J)$ be

$$\mathcal{L}^k(\alpha, \beta) \equiv \big(\mathcal{L}_1^k(\alpha, \beta), \dots, \mathcal{L}_J^k(\alpha, \beta)\big)'.$$

Let $\Gamma_k(\beta) \equiv \{\mathcal{L}^k(\alpha, \beta) : \alpha \in Y\}$ and $\breve{\Gamma}_k(\beta)$ be the convex hull of $\Gamma_k(\beta)$. By Lemma 3 of Chamberlain (1987), $\breve{\Gamma}_k(\beta) = \{\int \mathcal{L}^k(\alpha, \beta)\, dF(\alpha) : F$ is a CDF on $Y\}$. Therefore, $\int \mathcal{L}^k(\alpha, \beta)\, dF_k(\alpha) \in \breve{\Gamma}_k(\beta)$. Note that $\Gamma_k(\beta)$ is contained in the unit simplex and so has dimension $J-1$. By the Carathéodory theorem, there exist J vectors $\mathcal{L}^k(\alpha_m^k, \beta)$ $(m = 1, \dots, J)$ and $0 \le \pi_m^k \le 1$ with $\sum_{m=1}^J \pi_m^k = 1$ such that

$$\int \mathcal{L}^k(\alpha, \beta)\, dF_k(\alpha) = \sum_{m=1}^J \pi_m^k \mathcal{L}^k\big(\alpha_m^k, \beta\big),$$

giving the conclusion for the discrete distribution F_k^J with J support points at $(\alpha_1^k, \dots, \alpha_J^k)$ and probabilities $(\pi_1^k, \dots, \pi_J^k)$.

Next, for any $\epsilon > 0$, let $\beta \in B$ and $F_{k\beta} \in \mathcal{F}_k(\beta, \mathcal{P})$ satisfy

$$\Delta_u^k - \epsilon < \int \Delta(\alpha, \beta)\, dF_{k\beta}(\alpha) \equiv \overline{\Delta}(\beta).$$

Similarly to the previous paragraph, let $\Gamma_k^\Delta(\beta) \equiv \{(\mathcal{L}^k(\alpha, \beta)', \Delta(\alpha, \beta))' : \alpha \in Y\}$ and $\breve{\Gamma}_k^\Delta(\beta)$ be the convex hull of $\Gamma_k^\Delta(\beta)$. Then $(\mathcal{P}_1^k, \dots, \mathcal{P}_J^k, \overline{\Delta}(\beta))' \in \breve{\Gamma}_k^\Delta(\beta)$, so by Caratheodory's theorem, there exists a discrete distribution $F_{k\beta}^{J+1}$ with $J+1$ support points $(\alpha_1^k, \dots, \alpha_{J+1}^k)$ and probabilities $\pi_1^k, \dots, \pi_{J+1}^k$ such that $F_{k\beta}^{J+1} \in \mathcal{F}_k(\beta, \mathcal{P})$ and $\int \Delta(\alpha, \beta)\, dF_{k\beta}^{J+1}(\alpha) = \overline{\Delta}(\beta)$.

We now show that it suffices to have mass over just J points. Consider the problem of allocating $\pi_1^k, \dots, \pi_{J+1}^k$ among $(\alpha_1^k, \dots, \alpha_{J+1}^k)$ so as to solve

$$\max_{(\pi_1^k, \dots, \pi_{J+1}^k)} \sum_{m=1}^{J+1} \Delta\big(\alpha_m^k, \beta\big)\pi_m^k, \quad \text{s.t.}$$

$$\sum_{m=1}^{J+1} \pi_m^k \mathcal{L}_j^k\big(\alpha_m^k, \beta\big) = \mathcal{P}_j^k, \quad \sum_{m=1}^{J+1} \pi_m^k = 1, \pi_m^k \ge 0 \qquad (m = 1, \dots, J+1).$$

This is a linear program of the form

$$\max_{\pi^k \in \mathbb{R}^{J+1}} c'\pi^k \quad \text{such that} \quad \pi^k \ge 0, \quad A\pi^k = b, \quad 1'\pi^k = 1,$$

and any basic feasible solution to this program has $J+1$ active constraints, of which at most $\operatorname{rank}(A) + 1$ can be equality constraints. This means that at least $J+1-\text{rank}([A', 1]')$ of active constraints are of the form $\pi_m^k = 0$; see, for example, Theorem 2.3 and Definition 2.9(ii) in Bertsimas and Tsitsiklis (1997). Since each column of A sums to 1, $\text{rank}([A', 1]') \leq J$ and a basic solution to this linear programming problem will have at least one zero. Thus, there are at most J strictly positive π_m^k's.[1] Therefore, we have shown that there exists a distribution $F_{k\beta}^J \in \mathcal{F}_k(\beta, \mathcal{P})$ with just J points of support such that

$$\Delta_u^k - \epsilon < \int \Delta(\alpha, \beta)\, dF_{k\beta}^{J+1}(\alpha) \leq \int \Delta(\alpha, \beta)\, dF_{k\beta}^{J}(\alpha).$$

This construction works for every $\epsilon > 0$. *Q.E.D.*

S.8.2. *Numerical Results for Logit Model*

We carry out some additional numerical calculations for the logit model where

$$Y_{it} = 1\big(\beta^* X_{it} + \alpha_i \geq \varepsilon_{it}\big), \quad \varepsilon_{it} \sim L(0, 1),$$
$$X_{it} = 1(\alpha_i \geq \eta_{it}), \quad \eta_{it} \sim N(0, 1), \quad \alpha_i \sim N(0, 1),$$

where $L(0, 1)$ denotes the standard logistic distribution normalized to have zero mean and unit variance. We consider different DGPs indexed by $\beta^* \in [-2, 2]$ and $T \in \{2, 3\}$. Figures S.1 and S.2 show nonparametric bounds for ATEs and semiparametric bounds for β^* and ATEs for $T = 2$ and $T = 3$, respectively. The semiparametric bounds are obtained using the computational algorithm described in Section 8 of the paper with $M = 100$ and $\lambda_M = 1.3 \times 10^{-8}$. The elements of the fixed grid Y_M are located at the percentiles of the standard normal distribution. As is well known, we find that β^* is identified for $T \geq 2$. The nonparametric bounds for the ATEs (NP-bounds) can be very wide, even when we impose monotonicity (NPM-bounds). The semiparametric bounds for the ATEs (SP-bounds) are tighter than the nonparametric bounds and shrink exponentially fast with T, as shown in Theorem 6.

S.8.3. *Proof of Lemma 8*

Consider the set $\overline{\Re} = (-\infty, +\infty) \cup \{-\infty, +\infty\}$. By assumption, $H(v)$ is strictly monotonic and continuous on $\overline{\Re}$ with $H(-\infty) = 0$ and

[1]Note that $\text{rank}([A', 1]') \leq J$, since $\sum_{j=1}^J \mathcal{L}_j^k(\alpha, \beta) = 1$. The exact rank of $[A', 1]'$ depends on the sequence X^k, the parameter β, the form of $\mathcal{L}_j^k(\alpha, \beta)$, and T. For example, in the model of equation (8) of the main text with $T = 2$ and X binary, $\text{rank}(A) = J - 2 = 2$ when $x_1 = x_2$, $\beta = 0$, or H is the logistic distribution; whereas $\text{rank}(A) = J - 1 = 3$ for $X_1^k \neq X_2^k$, $\beta \neq 0$, and H is any continuous distribution different from the logistic.

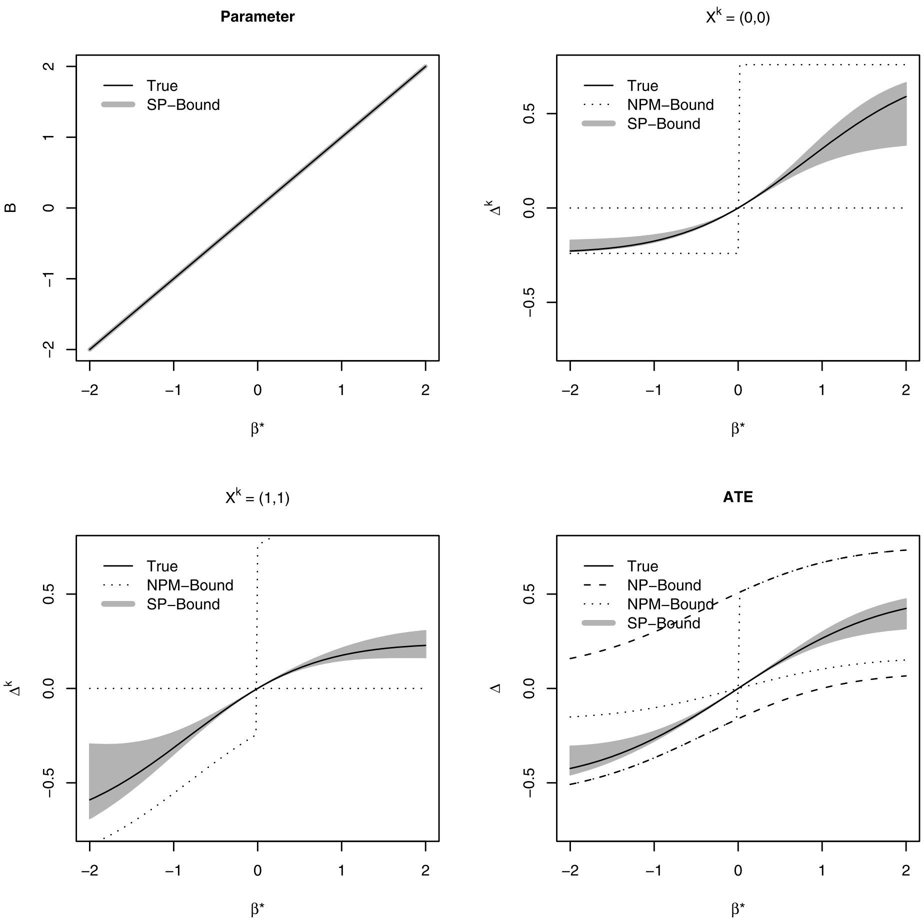

FIGURE S.1.—Identified set for parameter and ATEs in binary-choice logit models with $Y_{it} = 1(\beta^* X_{it} + \alpha_i \geq \varepsilon_{it})$, $\varepsilon_{it} \sim L(0,1)$, $X_{it} = 1(\alpha_i \geq \eta_{it})$, $\eta_{it} \sim N(0,1)$, $\alpha_i \sim N(0,1)$, $\beta^* \in [-2,2]$, and $T = 2$.

$H(+\infty) = 1$. Let $H^{-1}(u)$ be the inverse function defined on $[0,1]$. Let $\overline{v} = \max_{X^k \in \{X^1,\dots,X^K\}, \beta \in B} |X_t^{k\prime} \beta|$ and define the function

$$T(u) = \begin{cases} \overline{v} + H^{-1}(u), & \frac{3}{4} \leq u \leq 1, \\ (4u-2)\left[\overline{v} + H^{-1}\left(\frac{3}{4}\right)\right], & \frac{1}{4} < u < \frac{3}{4}, \\ -\overline{v} + H^{-1}(u), & 0 \leq u \leq \frac{1}{4}. \end{cases}$$

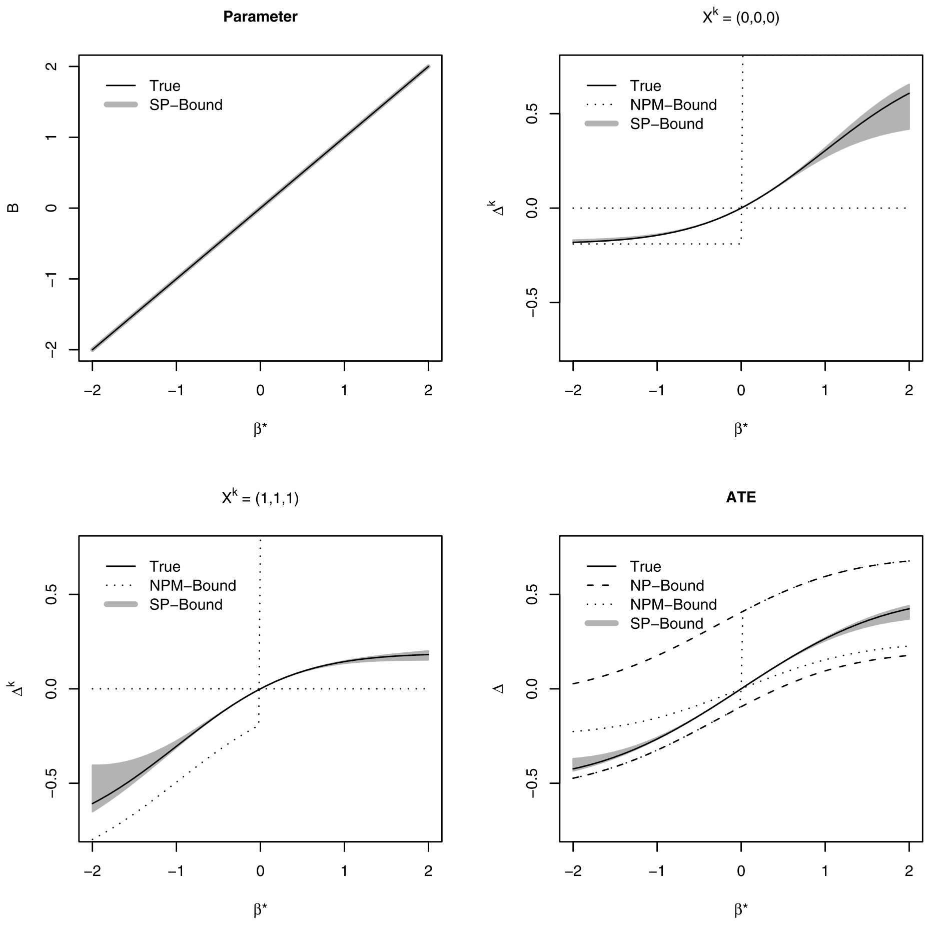

FIGURE S.2.—Identified set for parameter and ATEs in binary-choice logit models with $Y_{it} = 1(\beta^* X_{it} + \alpha_i \geq \varepsilon_{it})$, $\varepsilon_{it} \sim L(0,1)$, $X_{it} = 1(\alpha_i \geq \eta_{it})$, $\eta_{it} \sim N(0,1)$, $\alpha_i \sim N(0,1)$, $\beta^* \in [-2,2]$, and $T = 3$.

This function is continuous and differentiable except at $u = \frac{1}{4}$ and $u = \frac{3}{4}$. At $u = \frac{1}{4}$, the left derivative is $[h(H^{-1}(\frac{1}{4}))]^{-1}$ and the right derivative is $4[\overline{v} + H^{-1}(\frac{3}{4})]$.

Consider the function $H(v + T(u))$. By the chain rule, $H(v + T(u))$ is differentiable everywhere on $[-\overline{v}, \overline{v}] \times (\frac{1}{4}, \frac{3}{4})$, and right differentiable at $(v, \frac{1}{4})$ and left differentiable at $(v, \frac{3}{4})$ with derivative (right or left) equal to

$$h\big(v + T(u)\big)4\left[\overline{v} + H^{-1}\left(\frac{3}{4}\right)\right].$$

This derivative is uniformly bounded on $[-\overline{v}, \overline{v}] \times (\frac{1}{4}, \frac{3}{4})$ by h uniformly bounded. Also $H(v+T(u))$ is differentiable everywhere on $[-\overline{v}, \overline{v}] \times \{(\frac{3}{4}, \infty) \cup (-\infty, \frac{1}{4})\}$, right differentiable at $[-\overline{v}, \overline{v}] \times \{\frac{3}{4}\}$, and left differentiable at $[-\overline{v}, \overline{v}] \times \{\frac{1}{4}\}$. For $u \in [3/4, 1]$, the (right) derivative is

$$\begin{aligned}\frac{\partial}{\partial u} H\big(v+T(u)\big) &= H'\big(v+T(u)\big)T'(u) \\ &= \frac{h(v+\overline{v}+H^{-1}(u))}{h(H^{-1}(u))} \le \frac{h(H^{-1}(u))}{h(H^{-1}(u))} = 1,\end{aligned}$$

where the inequality holds by $\overline{v}+v \ge 0$ (implied by $v \ge -\overline{v}$) and by $H^{-1}(u) > 0$. It follows similarly that $\partial H(v + T(u))/\partial u$ is uniformly bounded by 1 on $[-\overline{v}, \overline{v}] \times [0, \frac{1}{4}]$. It follows that there is a constant C such that, for all $v \in [-\overline{v}, \overline{v}]$ and $u, \tilde{u} \in [0, 1]$,

$$\big|H\big(v+T(\tilde{u})\big) - H\big(v+T(u)\big)\big| \le C|\tilde{u} - u|.$$

Note that $T^{-1}(\alpha)$ is a strictly monotonic increasing function on $\overline{\Re}$. Define $d(\tilde{\alpha}, \alpha) = |T^{-1}(\tilde{\alpha}) - T^{-1}(\alpha)|$. Note that $d(\tilde{\alpha}, \alpha) \ge 0$ with equality if and only if $\tilde{\alpha} = \alpha$, and for any three points $\overline{\alpha}$, $\tilde{\alpha}$, and α, the triangle inequality implies

$$\begin{aligned}d(\tilde{\alpha}, \alpha) &= \big|T^{-1}(\tilde{\alpha}) - T^{-1}(\alpha)\big| \\ &\le \big|T^{-1}(\tilde{\alpha}) - T^{-1}(\overline{\alpha})\big| + \big|T^{-1}(\overline{\alpha}) - T^{-1}(\alpha)\big| \\ &= d(\tilde{\alpha}, \overline{\alpha}) + d(\overline{\alpha}, \alpha).\end{aligned}$$

Therefore $d(\tilde{\alpha}, \alpha)$ is a metric. Also, for $\tilde{u} = T^{-1}(\tilde{\alpha})$ and $u = T^{-1}(\alpha)$, we have

$$\sup_{v\in[-\overline{v},\overline{v}]} \big|H(v+\tilde{\alpha}) - H(v+\alpha)\big| \le C\big|T^{-1}(\tilde{\alpha}) - T^{-1}(\alpha)\big| = Cd(\tilde{\alpha}, \alpha).$$

Also, by $|X_t^{k\prime}\beta| \le \overline{v}$, and $0 \le H(X_t^{k\prime}\beta + \alpha) \le 1$, for all t, k, and $\beta \in \mathbb{B}$,

$$\begin{aligned}&\big|\mathcal{L}_j^k(\tilde{\alpha}, \tilde{\beta}) - \mathcal{L}_j^k(\alpha, \beta)\big| \\ &\quad\le \big|\mathcal{L}_j^k(\tilde{\alpha}, \tilde{\beta}) - \mathcal{L}_j^k(\alpha, \tilde{\beta})\big| + \big|\mathcal{L}_j^k(\alpha, \tilde{\beta}) - \mathcal{L}_j^k(\alpha, \beta)\big| \\ &\quad\le Cd(\tilde{\alpha}, \alpha) + \sup_{\alpha,t,k}\big|H\big(X_t^{k\prime}\tilde{\beta} + \alpha\big) - H\big(X_t^{k\prime}\beta + \alpha\big)\big| \\ &\quad\le Cd(\tilde{\alpha}, \alpha) + \sup_{v} h(v) \sup_{t,k}\big\|X_t^k\big\| \|\tilde{\beta} - \beta\| \\ &\quad\le C\big[d(\tilde{\alpha}, \alpha) + \|\tilde{\beta} - \beta\|\big].\end{aligned}$$

Finally, for every M, let $\overline{\alpha}_{mM} = T((m-1)/(M-1))$ $(m = 1, \ldots, M)$. Then

$$\eta(M) = \sup_{\alpha \in \overline{\Re}} \min_{\tilde{\alpha} \in Y_M} d(\alpha, \tilde{\alpha}) = \sup_{u \in [0,1]} \min_{\tilde{u} \in \{0, 1/(M-1), 2/(M-1), \ldots, 1\}} |u - \tilde{u}| = 1/(M-1). \qquad \textit{Q.E.D.}$$

S.8.4. *Proof of Theorem 9*

This proof is omitted because it is very similar to (but easier than) the proof of Theorem 10 to follow.

S.9. SUPPLEMENTS TO SECTION 9

Here we describe the estimation algorithm, give the proofs of Theorems 10 and 11, and present an alternative inference method based on projection.

S.9.1. *Estimation: Implementation Details*

To implement the estimation method, we also start from simpler estimates of the bounds corresponding to those described in the computation section. Specifically, for $\hat{\pi}(\beta) \in \arg\min_{\pi \in S_M^K} \hat{T}_\lambda(\beta, \pi)$, let $\hat{S}^k(\beta) = \{\pi^k : P_j^k(\beta, \pi, \hat{M}) = P_j^k(\beta, \hat{\pi}(\beta), \hat{M}), j = 1, \ldots, J\}$ and let

$$\check{\Delta}_\ell^k = \min_{\beta \in \hat{B}, \pi^k \in \hat{S}^k(\beta)} \sum_{m=1}^{M} \pi_m^k \Delta(\overline{\alpha}_{mM}, \beta),$$

$$\check{\Delta}_u^k = \max_{\beta \in \hat{B}, \pi^k \in \hat{S}^k(\beta)} \sum_{m=1}^{M} \pi_m^k \Delta(\overline{\alpha}_{mM}, \beta).$$

We use these estimated bounds as starting values and then search over other possible values of π, similarly to the computational approach.

The choice of $\hat{M}$ is important for this estimator. In our empirical examples, we have proceeded by starting with a small $\hat{M}$, and stopping when the change in the estimated sets is small. We have found that quite small $\hat{M}$ often suffices. The choice of weights $\hat{w}_j^k$ is also important. The optimal choice, corresponding to minimum chi-square, would be $\hat{w}_j^k = \mathcal{P}^k / \mathcal{P}_j^k$. Using sample frequencies in place of population frequencies does not work well due to small cell sizes. One could use a two-step procedure where one first computes the identified set for weights like $\hat{w}_j^k = \hat{P}^k$ and then reestimates the identified set using weights $\hat{w}_j^k = \hat{P}_k / P_j^k(\beta, \hat{\pi}(\beta), \hat{M})$ for some $\beta \in \hat{B}$.

S.9.2. *Proof of Theorem 10*

For notational convenience, we here denote the probabilities associated with the fixed grid $\{\overline{\alpha}_{1M}, \dots, \overline{\alpha}_{MM}\}$ by $\overline{\pi}^k$. Let $\overline{\pi} = (\overline{\pi}^{1\prime}, \dots, \overline{\pi}^{K\prime})'$ be a $KM \times 1$ vector with each $\overline{\pi}^k$ in the M-dimensional unit simplex $\mathcal{S}_M$. Also, let the probabilities associated with a variable grid $\{\alpha_1^k, \dots, \alpha_{J+1}^k\}$ be π^k so that $\pi = (\pi^{1\prime}, \dots, \pi^{K\prime})'$ is a $[(J+1)K] \times 1$ vector of probabilities with each π^k in the $J+1$-dimensional unit simplex $\mathcal{S}_{J+1}$. Let $\alpha^k = (\alpha_1^k, \dots, \alpha_{J+1}^k)'$, $\alpha = (\alpha^{1\prime}, \dots, \alpha^{K\prime})'$, $\gamma = (\alpha', \pi')'$, $\theta = (\beta', \gamma')'$, $\tilde{P}_j^k(\theta) = \sum_{\ell=1}^{J+1} \mathcal{L}_j^k(\alpha_\ell^k, \beta)\pi_\ell^k$, $\Delta^k(\theta) = \sum_{\ell=1}^{J+1} \Delta(\alpha_\ell^k, \beta)\pi_\ell^k$, $\Theta = \mathbb{B} \times Y^{(J+1)K} \times \mathcal{S}_{J+1}^K$, and

$$\hat{Q}(\theta) = \sum_{j,k} \hat{w}_j^k \big[\hat{P}_j^k - \tilde{P}_j^k(\theta)\big]^2, \quad Q(\theta) = \sum_{j,k} w_j^k \big[\mathcal{P}_j^k - \tilde{P}_j^k(\theta)\big]^2.$$

By applying the Caratheodory theorem as in the proof of Lemma 12, for every $\overline{\pi}$ there is $\theta(\overline{\pi}, \beta) = (\beta', \gamma(\overline{\pi}, \beta)')'$ with

$$\Delta^k\big(\theta(\overline{\pi}, \beta)\big) = \sum_{m=1}^{M} \Delta(\overline{\alpha}_{mM}, \beta)\overline{\pi}_m^k,$$
$$\tilde{P}_j^k\big(\theta(\overline{\pi}, \beta)\big) = P_j^k(\beta, \overline{\pi}, M) \qquad (j = 1, \dots, J; k = 1, \dots, K).$$

Let $\Theta_I = \{\theta : Q(\theta) = 0\}$,

$$\tilde{\Theta} = \big\{\theta(\overline{\pi}, \beta) : \hat{Q}\big(\theta(\overline{\pi}, \beta)\big) + \lambda_n \overline{\pi}'\overline{\pi} \le \epsilon_n\big\},$$
$$\Theta_M = \big\{\theta(\overline{\pi}, \beta) : \overline{\pi} \in \mathcal{S}_M^K, \beta \in \mathbb{B}\big\}.$$

By construction, the projection of $\tilde{\Theta}$ on $\mathbb{B}$ coincides with $\hat{B}$ and the projection of Θ_I on $\mathbb{B}$ coincides with B. Also the identified set of marginal effects is $\{\Delta^k(\theta) : \theta \in \Theta_I\}$, $\Delta^k(\theta)$ is a continuous function of θ, and $\hat{D}^k = \{\Delta^k(\theta) : \theta \in \tilde{\Theta}\}$. Since the minimum and maximum of a set are continuous in the Hausdorff metric, it suffices to show that $d_H(\tilde{\Theta}, \Theta_I) \xrightarrow{p} 0$.

Let $d(\theta, \tilde{\theta}) = \max_{j,k} \max\{d(\alpha_j^k, \tilde{\alpha}_j^k), |\pi_j^k - \tilde{\pi}_j^k|, \|\beta - \tilde{\beta}\|\}$. From Assumption 9 and $\hat{M} \xrightarrow{p} \infty$, we have

$$\sup_{\alpha \in Y} \min_{\tilde{\alpha} \in Y_{\hat{M}}} d(\alpha, \tilde{\alpha}) \le \eta(\hat{M}) \xrightarrow{p} 0.$$

Therefore, for every $\alpha \in Y$, there is $\overline{\alpha}_{m(\alpha),\hat{M}}$ with $d(\alpha, \overline{\alpha}_{m(\alpha),\hat{M}}) \le \eta(\hat{M})$, so that, for any $\theta \in \Theta$, there are $\overline{\alpha}_{m(\alpha_\ell^k),\hat{M}}$ with $\max_{1 \le \ell \le J+1,k}\{d(\alpha_\ell^k, \overline{\alpha}_{m(\alpha_\ell^k),\hat{M}})\} \le \eta(\hat{M})$. Let $\alpha^k(\theta) = (\overline{\alpha}_{m(\alpha_1^k),\hat{M}}, \dots, \overline{\alpha}_{m(\alpha_{J+1}^k),\hat{M}})'$, $\alpha(\theta) = (\alpha^1(\theta)', \dots, \alpha^K(\theta)')'$, and $\overline{\theta}(\theta) = (\beta', \alpha(\theta)', \pi')'$. By construction, $\overline{\theta}(\theta) \in \Theta_M$ and $d(\overline{\theta}(\theta), \theta) \le$

$\eta(\hat{M})$. Thus,

$$\sup_{\theta\in\Theta}\inf_{\tilde{\theta}\in\Theta_{\hat{M}}} d(\theta,\tilde{\theta}) \leq \eta(\hat{M}).$$

Also, by Assumption 9,

$$\left|\tilde{P}_j^k(\theta) - \tilde{P}_j^k(\tilde{\theta})\right| \leq \sum_{\ell=1}^{J}\left|\mathcal{L}_j^k\left(\alpha_\ell^k,\beta\right)\pi_\ell^k - \mathcal{L}_j^k\left(\tilde{\alpha}_\ell^k,\tilde{\beta}\right)\tilde{\pi}_\ell^k\right| \leq Cd(\theta,\tilde{\theta}).$$

It then follows by standard calculations that there is $\hat{C} = O_p(1)$ such that

$$\left|\hat{Q}(\theta) - \hat{Q}(\tilde{\theta})\right| \leq \hat{C}d(\theta,\tilde{\theta}) \quad \text{for all } \theta,\tilde{\theta}\in\Theta.$$

Therefore we have

$$\sup_{\theta\in\Theta}\inf_{\tilde{\theta}\in\Theta_{\hat{M}}}\left|\hat{Q}(\theta) - \hat{Q}(\tilde{\theta})\right| \leq \hat{C}\eta(\hat{M}).$$

Also note that

$$\sup_{\theta\in\Theta_I}\hat{Q}(\theta) = \sum_{j,k}\hat{w}_j^k\left[\hat{P}_j^k - \mathcal{P}_j^k\right]^2 = O_p\left(n^{-1}\right).$$

Next let $\delta > 0$ be any positive constant and define the events

$$\mathcal{E}_1 = \left\{\eta(\hat{M}) < \delta\right\}, \quad \mathcal{E}_2 = \left\{\hat{C}\eta(\hat{M}) < \frac{\epsilon_n}{3}\right\},$$
$$\mathcal{E}_3 = \left\{\sup_{\theta\in\Theta_I}\hat{Q}(\theta) < \frac{\epsilon_n}{3}\right\}, \quad \mathcal{E}_4 = \sup_{\overline{\pi}\in\mathcal{S}_M^K}\lambda_n\overline{\pi}'\overline{\pi} < \frac{\epsilon_n}{3}.$$

By $(n^{-1} + \eta(\hat{M}) + \lambda_n)/\epsilon_n \xrightarrow{p} 0$, it follows that

$$\Pr(\mathcal{E}_1) \longrightarrow 1,$$
$$\Pr(\mathcal{E}_2) = \Pr\left(\hat{C} < \frac{\eta(\hat{M})^{-1}\epsilon_n}{3}\right) \longrightarrow 1,$$
$$\Pr(\mathcal{E}_3) = \Pr\left(n\sup_{\theta\in\Theta_I}\hat{Q}(\theta) < \frac{n\epsilon_n}{3}\right) \longrightarrow 1,$$
$$\Pr(\mathcal{E}_4) \geq \Pr\left(\lambda_n K \leq \frac{\epsilon_n}{3}\right) \longrightarrow 1.$$

It follows that $\Pr(\bigcap_{r=1}^{4}\mathcal{E}_r) \longrightarrow 1$. When $\bigcap_{r=1}^{4}\mathcal{E}_r$ occurs, then, for every $\theta \in \Theta_I$, there is $\overline{\pi}$ with $\theta_M = \theta(\overline{\pi}, \beta) \in \Theta_M$ such that $d(\theta, \overline{\theta}) < \delta$ and

$$\begin{aligned}\hat{Q}(\overline{\theta}) + \lambda_n \overline{\pi}'\overline{\pi} &\le \hat{Q}(\overline{\theta}) + \frac{\epsilon_n}{3} \le \hat{Q}(\theta) + \hat{Q}(\overline{\theta}) - \hat{Q}(\theta) + \frac{\epsilon_n}{3} \\ &\le \sup_{\theta\in\Theta_I} \hat{Q}(\theta) + \hat{C}\hat{\eta}(M) + \frac{\epsilon_n}{3} \le \epsilon_n,\end{aligned}$$

that is, $\overline{\theta} \in \tilde{\Theta}$. Thus, with probability approaching 1,

$$\sup_{\theta\in\Theta_I} \inf_{\tilde{\theta}\in\tilde{\Theta}} d(\theta, \tilde{\theta}) \le \delta.$$

Next, note that $\hat{Q}(\theta) \xrightarrow{p} Q(\theta)$, so it follows by Theorem 2.1 of Newey (1991) that $\sup_{\theta\in\Theta} |\hat{Q}(\theta) - Q(\theta)| \xrightarrow{p} 0$. Define $\Theta_I^\delta = \{\theta : \inf_{\tilde{\theta}\in\Theta_I} d(\theta, \tilde{\theta}) < \delta\}$. Note that Θ_I^δ is open so that $\Theta \setminus \Theta_I^\delta$ is compact, so by continuity of $Q(\theta)$, $\inf_{\Theta\setminus\Theta_I^\delta} Q(\theta) = \rho > 0$. It follows by uniform convergence that $\inf_{\Theta\setminus\Theta_I^\delta} \hat{Q}(\theta) > \frac{\rho}{2}$ with probability approaching 1 (w.p.a.1). By $\epsilon_n \to 0$,

$$\sup_{\theta\in\tilde{\Theta}} \hat{Q}(\theta) \le \sup_{\overline{\pi}} \bigl\{\hat{Q}\bigl(\theta(\overline{\pi}, \beta)\bigr) + \lambda_n \overline{\pi}'\overline{\pi} \le \epsilon_n\bigr\} < \rho/2,$$

so that $\tilde{\Theta} \subseteq \Theta_I^\delta$. Therefore w.p.a.1, for all $\tilde{\theta} \in \tilde{\Theta}$. there exists $\theta \in \Theta_I$ such that $d(\tilde{\theta}, \theta) < \delta$, that is, $\sup_{\tilde{\theta}\in\tilde{\Theta}} \inf_{\theta\in\Theta_I} d(\theta, \tilde{\theta}) \le \delta$. It follows that with w.p.a.1, $d_H(\tilde{\Theta}, \Theta_I) \le \delta$. Since $\delta > 0$ is arbitrary, it follows that $d_H(\tilde{\Theta}, \Theta_I) \xrightarrow{p} 0$. *Q.E.D.*

S.9.3. *Proof of Theorem 11*

We have that, for $S_n(\mathcal{P}) = \hat{\theta} - \theta^* = \hat{\theta} - \theta^*(\mathcal{P})$,

$$\begin{aligned}&\Pr_\Pi\bigl\{\theta^* \notin [\underline{\theta}, \overline{\theta}]\bigr\} \\ &\quad= \Pr_\Pi\bigl\{S_n(\mathcal{P}) \notin \bigl[\underline{G}_n^{-1}(\alpha_2, \mathcal{P}), \overline{G}_n^{-1}(1-\alpha_1, \mathcal{P})\bigr]\bigr\} \\ &\quad\le \Pr_\Pi\bigl(\bigl\{S_n(\mathcal{P}) \notin \bigl[\underline{G}_n^{-1}(\alpha_2, \mathcal{P}), \overline{G}_n^{-1}(1-\alpha_1, \mathcal{P})\bigr]\bigr\} \\ &\qquad\cap \bigl\{\mathcal{P} \in \mathrm{CR}_{1-\gamma}(\mathcal{P})\bigr\}\bigr) + \Pr_\Pi\bigl\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\bigr\} \\ &\quad\le \Pr_\Pi\bigl(\bigl\{S_n(\mathcal{P}) \notin \bigl[G_n^{-1}(\alpha_2, \mathcal{P}), G_n^{-1}(1-\alpha_1, \mathcal{P})\bigr]\bigr\} \\ &\qquad\cap \bigl\{\mathcal{P} \in \mathrm{CR}_{1-\gamma}(\mathcal{P})\bigr\}\bigr) + \Pr_\Pi\bigl\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\bigr\} \\ &\quad\le \Pr_\Pi\bigl\{S_n(\mathcal{P}) \notin \bigl[G_n^{-1}(\alpha_2, \mathcal{P}), G_n^{-1}(1-\alpha_1, \mathcal{P})\bigr]\bigr\} \\ &\qquad+ \Pr_\Pi\bigl\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\bigr\} \\ &\quad\le \alpha + \Pr_\Pi\bigl\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\bigr\}.\end{aligned}$$

Thus if $\limsup_{n\to\infty} \Pr_\Pi\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\} \leq \gamma$, we obtain that $\lim_n \Pr_\Pi\{\theta^* \notin [\underline{\theta}, \overline{\theta}]\} \leq \alpha + \gamma$, which is the desired conclusion.

It now remains to show that $\limsup_{n\to\infty} \Pr_\Pi\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\} \leq \gamma$. We have that

$$\Pr_\Pi\{\mathcal{P} \notin \mathrm{CR}_{1-\gamma}(\mathcal{P})\} = \Pr_\Pi\{W(\mathcal{P}, P) > c_{1-\gamma}(\chi^2_{K(J-1)})\}.$$

By the uniform central limit theorem, $W(\mathcal{P}, \hat{P})$ converges in law to $\chi^2_{K(J-1)}$ under any sequence Π in $\mathbb{P}$. Therefore,

$$\lim_{n\to\infty} \Pr_\Pi\{W(\mathcal{P}, \hat{P}) > c_{1-\gamma}(\chi^2_{K(J-1)})\} = \Pr\{\chi^2_{K(J-1)} > c_{1-\gamma}(\chi^2_{K(J-1)})\} = \gamma.$$

Q.E.D.

S.9.4. *Modified Projection Method*

The following method projects a confidence region for conditional choice probabilities onto a simultaneous confidence region for all possible ATEs and other structural parameters. In general, this method is more conservative than the perturbed bootstrap method when a single ATE or structural parameter is of interest. We include a more detailed comparison between the two methods at the end of this section.

It is convenient to describe the modified projection method in two stages.

Stage 1. The probabilities $\mathcal{P}_j^k$ belong to the product $\mathcal{S}_J^K$ of K unit simplexes of dimension J. We can begin by constructing a confidence region for the true choice probabilities $\mathcal{P}$ by collecting all probabilities $P = (P_1^1, \dots, P_J^1, \dots, P_J^K)' \in \mathcal{S}_J^K$ that pass a goodness-of-fit test:

$$\mathrm{CR}_{1-\alpha}(\mathcal{P}) = \{P \in \mathcal{S}_J^K : W(P, \hat{P}) \leq c_{1-\alpha}(\chi^2_{K(J-1)})\},$$

where $c_{1-\alpha}(\chi^2_{K(J-1)})$ is the $(1-\alpha)$-quantile of the $\chi^2_{K(J-1)}$ distribution and W is the goodness-of-fit statistic:

$$W(P, \hat{P}) = n \sum_{j,k} \hat{P}^k \frac{(\hat{P}_j^k - P_j^k)^2}{P_j^k}.$$

Stage 2. To construct confidence regions for marginal effects and any other structural parameters, we project each $P \in \mathrm{CR}_{1-\alpha}(\mathcal{P})$ onto $\Xi = \{P : \exists \beta \in \mathbb{B}$ with $\mathcal{F}_k(\beta, P) \neq \emptyset, \forall k = 1, \dots, K\}$, the space of conditional choice probabilities that is compatible with the model. We obtain this projection $P^*(P)$ by solving the minimum distance problem:

$$P^*(P) = \arg\min_{\tilde{P}\in\Xi} W(\tilde{P}, P), \quad W(\tilde{P}, P) = n \sum_{j,k} \hat{P}^k \frac{(P_j^k - \tilde{P}_j^k)^2}{\tilde{P}_j^k}.$$

The confidence regions are then constructed from the projections of all the choice probabilities in $\mathrm{CR}_{1-\alpha}(\mathcal{P})$. For the identified set of the model parameter, for example, for each $P \in \mathrm{CR}_{1-\alpha}(\mathcal{P})$, we solve

$$B^*(P) = \{\beta \in \mathbb{B} : \exists \tilde{P} \in P^*(P) \text{ with } \mathcal{F}_k(\beta, \tilde{P}) \neq \emptyset, k = 1, \dots, K\}.$$

Denote the resulting confidence region as

$$\mathrm{CR}_{1-\alpha}(B^*) = \{B^*(P) : P \in \mathrm{CR}_{1-\alpha}(\mathcal{P})\}.$$

We may interpret this set as a confidence region for the set B^* of β that are compatible with a best approximating model. Under correct specification, this will be a confidence region for the identified set B.

If we are interested in bounds on marginal effects, for each $P \in \mathrm{CR}_{1-\alpha}(\mathcal{P})$ we get

$$\Delta_\ell^k(P) = \min_{\beta \in B^*(P), F_k \in \mathcal{F}_k(\beta, P^*(P))} \int \Delta(\alpha, \beta)\, dF_k(\alpha),$$

$$\Delta_u^k(P) = \max_{\beta \in B^*(P), F_k \in \mathcal{F}_k(\beta, P^*(P))} \int \Delta(\alpha, \beta)\, dF_k(\alpha).$$

Denote the resulting confidence regions as

$$\mathrm{CR}_{1-\alpha}[\Delta_\ell^{k*}, \Delta_u^{k*}] = \{[\Delta_\ell^k(P), \Delta_u^k(P)] : P \in \mathrm{CR}_{1-\alpha}(\mathcal{P})\}.$$

These sets are confidence regions for the sets $[\Delta_\ell^{k*}, \Delta_u^{k*}]$, where Δ_ℓ^{k*} and Δ_u^{k*} are the lower and upper bounds on the marginal effects induced by any best-approximating model. Under correct specification, these will include the true upper and lower bounds on the marginal effect $[\Delta_\ell^k, \Delta_u^k]$ induced by any true model in $(B, \mathcal{P})$.

In a canonical projection method, we would implement the second stage by simply intersecting $\mathrm{CR}_{1-\alpha}(\mathcal{P})$ with Ξ, but this may give an empty intersection either in finite samples or under misspecification. We avoid this problem by using the projection step instead of the intersection, and also by retargeting our confidence regions onto the best approximating model.

THEOREM A15: *If Assumptions* 5, 8, *and* 9 *are satisfied*, *then*, *for any sequence of data-generating process* $\Pi = \Pi_n$ *satisfying Assumption* 10,

$$\lim_{n\to\infty} \Pr_\Pi\big[\{\mathcal{P} \in \mathrm{CR}_{1-\alpha}(\mathcal{P})\} \cap \{B^* \in \mathrm{CR}_{1-\alpha}(B^*)\} \cap \{[\Delta_\ell^{k*}, \Delta_u^{k*}] \in \mathrm{CR}_{1-\alpha}[\Delta_\ell^{k*}, \Delta_u^{k*}], \forall k\}\big] = 1 - \alpha.$$

PROOF: By the uniform central limit theorem, $W(\mathcal{P}, \hat{P})$ converges in law to $\chi^2_{J(K-1)}$ under any sequence of true DGPs with Π in $\mathbb{P}$. It follows that

$$\lim_{n\to\infty} \Pr_\Pi\{\mathcal{P} \in \mathrm{CR}_{1-\alpha}(\mathcal{P})\} = 1 - \alpha.$$

Further, the event $\mathcal{P} \in \mathrm{CR}_{1-\alpha}(\mathcal{P})$ implies then the event $P^*(\mathcal{P}) \in \{P^*(P) : P \in \mathrm{CR}_{1-\alpha}(\mathcal{P})\}$ by construction, which in turn implies the events $B^* \in \mathrm{CR}_{1-\alpha}(B^*)$ and $[\Delta_\ell^{k*}, \Delta_u^{k*}] \in \mathrm{CR}_{1-\alpha}[\Delta_\ell^{k*}, \Delta_u^{k*}]$, $\forall k$. *Q.E.D.*

We conclude by giving a comparison of the modified projection and perturbed bootstrap methods. The modified projection method is well suited for performing simultaneous inference on all possible functionals of the parameter vector. In contrast, the perturbed bootstrap is better suited for performing inference on a given functional of the parameter vector, such as the average structural effect. To understand why the latter method can be much sharper than the former method in the case where a single functional is of interest, it suffices to think of how these methods perform in the simplest situation of inference about the mean of a multinomial distribution. In this case, the perturbed bootstrap will become asymptotically equivalent to the usual bootstrap, since the limit distribution is continuous with respect to the DGP in this example, and our local perturbations of DGP converge to the true DGP (note that, more generally, in cases with limit distributions being discontinuous with respect to the DGP, the introduction of the local perturbations ensures that the resulting confidence interval possesses locally uniform coverage). Therefore, in this example, perturbed bootstrap inference asymptotically becomes first-order equivalent to the t-statistic-based inference on the mean, and is efficient. Now compare that with the Scheffé-style projection based confidence interval, whereby one creates a confidence region for multinomial probabilities and projects it down to the confidence interval for the mean, a linear functional of these probabilities. It is clear that the latter is very conservative, and is much less sharp than the t-statistic based confidence interval. We refer the reader to Romano and Wolf (2000) for the pertinent discussion of this example in the context of a closely related inference method.

REFERENCES

BERTSIMAS, D., AND J. N. TSITSIKLIS (1997): *Introduction to Linear Optimization*. Belmont, MA: Athena Scientific. [28]

CHAMBERLAIN, G. (1982): "Multivariate Regression Models for Panel Data," *Journal of Econometrics*, 18, 5–46. [1,3,4,19]

——— (1985): "Heterogeneity, Omitted Variables Bias, and Duration Dependence," in *Longitudinal Analysis of Labor Market Data*, ed. by J. J. Heckman and S. Singer. Cambridge: Cambridge University Press. [24]

——— (1987): "Asymptotic Efficiency in Estimation With Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334. [27]

COX, D. R. (1958): "The Regression Analysis of Binary Sequences," *Journal of the Royal Statistical Society, Ser. B*, 20, 215–232. [24]

FERNÁNDEZ-VAL, I. (2009): "Fixed Effects Estimation of Structural Parameters and Marginal Effects in Panel Probit Models," *Journal of Econometrics*, 150 (1), 71–85. [25]

HAHN, J., AND W. NEWEY (2004): "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica*, 72, 1295–1319. [25]

JUDD, K. L. (1998): *Numerical Methods in Economics*. Cambridge, MA: MIT Press. [24]

NEWEY, W. K. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity," *Econometrica*, 59, 1161–1167. [35]

ROMANO, J. P., AND M. WOLF (2000): "Finite Sample Nonparametric Inference and Large Sample Efficiency," *The Annals of Statistics*, 28 (3), 756–778. [38]

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, U.S.A.; vchern@mit.edu,

Dept. of Economics, Boston University, Boston, MA 02215, U.S.A.; ivanf@bu.edu,

Dept. of Economics, University of California, Los Angeles, Los Angeles, CA 90095, U.S.A.; hahn@econ.ucla.edu,

and

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, U.S.A.; wnewey@mit.edu.

Manuscript received February, 2009; final revision received February, 2012.