Supplement to: Binary Response Panel Data Models with Sample Selection and Self Selection

by Anastasia Semykina and Jeffrey M. Wooldridge

1. Asymptotic Variance of the Parametric Two-Step Estimator of Censored Selection Model

This Section discusses the estimation of the asymptotic variance of the estimator summarized in Procedure 3.2. Denote $\hat{\mathbf{w}}_{it} = (1, \mathbf{x}_{it}, \bar{\mathbf{z}}_i, \hat{v}_{it2})$, $\boldsymbol{\theta} = (\eta_1, \boldsymbol{\beta}', \boldsymbol{\xi}_1', \gamma)'$, $\mathbf{q}_{it} = (1, \mathbf{z}_{it}, \bar{\mathbf{z}}_i)$, and $\boldsymbol{\pi} = (\eta_2, \boldsymbol{\delta}', \boldsymbol{\xi}_2')'$. Using the argument similar to that presented in Section 2 below, it can be shown that

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \sim Normal(0, \mathbf{V}),$$

where $\mathbf{V} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ is the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$,

$$\mathbf{A} = \mathrm{E}\left[-\sum_{t=1}^{T} 1[s_{it} > 0] \cdot \mathbb{H}_{it}(\boldsymbol{\theta})\right], \tag{1}$$

$$\mathbf{B} = \mathrm{E}\left[\mathbf{p}_i \mathbf{p}_i',\right]$$

$$\mathbf{p}_i = \sum_{t=1}^{T} 1[s_{it} > 0] \cdot \left\{ \frac{\phi(\mathbf{w}_{it}\boldsymbol{\theta})}{\Phi(\mathbf{w}_{it}\boldsymbol{\theta})[1 - \Phi(\mathbf{w}_{it}\boldsymbol{\theta})]} \mathbf{w}_{it}'[y_{it} - \Phi(\mathbf{w}_{it}\boldsymbol{\theta})] - \mathbf{F}\mathbf{r}_i(\boldsymbol{\pi}) \right\},$$

$$\mathbf{F} = -\mathrm{E}\left[\sum_{t=1}^{T} \frac{1[s_{it} > 0] \cdot \phi(\mathbf{w}_{it}\boldsymbol{\theta})^2}{[1 - \Phi(\mathbf{w}_{it}\boldsymbol{\theta})]} \mathbf{w}_{it}'\mathbf{q}_{it}\gamma\right],$$

$$\mathbf{r}_i(\boldsymbol{\pi}) = \left[\mathrm{E}\left(-\sum_{t=1}^{T} \mathbf{H}_{it}(\boldsymbol{\pi})\right)\right]^{-1} \sum_{t=1}^{T} \mathbf{S}_{it}(\boldsymbol{\pi}). \tag{2}$$

Here $\mathbb{H}_{it}(\boldsymbol{\theta})$ is the Hessian matrix from the second-step probit estimation, while $\mathbf{H}_{it}(\boldsymbol{\pi})$ and $\mathbf{S}_{it}(\boldsymbol{\pi})$ are the Hessian matrix and score vector from the first-step Tobit estimation, respectively.

Then, $\text{Avar}(\hat{\boldsymbol{\theta}})$ can be estimated as $\hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}\hat{\mathbf{A}}^{-1}/N$,

$$\hat{\mathbf{A}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} 1[s_{it} > 0] \cdot \mathbb{H}_{it}(\hat{\boldsymbol{\theta}}), \tag{3}$$

$$\hat{\mathbf{B}} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbf{p}}_i \hat{\mathbf{p}}_i',$$

$$\hat{\mathbf{p}}_i = \sum_{t=1}^{T} 1[s_{it} > 0] \cdot \left\{ \frac{\phi(\hat{\mathbf{w}}_{it}\hat{\boldsymbol{\theta}})}{\Phi(\hat{\mathbf{w}}_{it}\hat{\boldsymbol{\theta}})[1-\Phi(\hat{\mathbf{w}}_{it}\hat{\boldsymbol{\theta}})]}\hat{\mathbf{w}}_{it}'[y_{it} - \Phi(\hat{\mathbf{w}}_{it}\hat{\boldsymbol{\theta}})] - \hat{\mathbf{F}}\mathbf{r}_i(\hat{\boldsymbol{\pi}}) \right\},$$

$$\hat{\mathbf{F}} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \left[\frac{1[s_{it} > 0] \cdot \phi(\hat{\mathbf{w}}_{it}\hat{\boldsymbol{\theta}})^2}{[1-\Phi(\hat{\mathbf{w}}_{it}\hat{\boldsymbol{\theta}})]} \hat{\mathbf{w}}_{it}'\mathbf{q}_{it}\hat{\gamma} \right],$$

$$\mathbf{r}_i(\hat{\boldsymbol{\pi}}) = \left[-\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{H}_{it}(\hat{\boldsymbol{\pi}}) \right]^{-1} \sum_{t=1}^{T} \mathbf{S}_{it}(\hat{\boldsymbol{\pi}}), \tag{4}$$

where $\mathbb{H}_{it}(\hat{\boldsymbol{\theta}})$ is the Hessian matrix from the second-step probit estimation evaluated at $\hat{\boldsymbol{\theta}}$, while $\mathbf{H}_{it}(\hat{\boldsymbol{\pi}})$ and $\mathbf{S}_{it}(\hat{\boldsymbol{\pi}})$ are the Hessian matrix and score vector from the first-step Tobit estimation, respectively, evaluated at $\hat{\boldsymbol{\pi}}$.

2. Asymptotic Properities of the Semiparametric Estimator

In this Section we discuss asymptotic properties of the semiparametric estimator proposed in Section 4 of the paper. The argument below is very similar to the one in Blundell and Powell (2004).

To demonstrate the consistency of the semiparametric estimator, first show that $\hat{\mathbf{S}}^t$ is consistent for $\boldsymbol{\Sigma}_0^t$, $t = 1, ..., T$, where $\boldsymbol{\Sigma}_0^t$ is a particular form of matrix $\boldsymbol{\Sigma}_\omega^t$ that uses the weighting matrix specified in equation (55) in the paper. Using the first-order mean-value expansion, for each t we can write:

$$\hat{\mathbf{S}}^t = \mathbf{S}_0^t + \mathbf{S}_1^t, \quad \text{where} \tag{5}$$

$$\mathbf{S}_l^t \equiv \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum\nolimits_{i<j} \omega_{ijt}^l (\mathbf{w}_{it} - \mathbf{w}_{jt})'(\mathbf{w}_{it} - \mathbf{w}_{jt}), \; l = 0, 1, \tag{6}$$

$$\omega_{ijt}^0 \equiv \frac{1}{h_\omega^2} \kappa_g \left(\frac{g_{it} - g_{jt}}{h_\omega} \right) \kappa_v \left(\frac{v_{it2} - v_{jt2}}{h_\omega} \right) d_{it} \cdot d_{jt} \cdot \tau_{it} \cdot \tau_{jt}, \tag{7}$$

$$\begin{aligned} \omega_{ijt}^1 \equiv & \frac{1}{h_\omega^3} \left\{ \kappa_g^{(1)} \left(\frac{g_{ijt}^*}{h_\omega} \right) \kappa_v \left(\frac{v_{ijt2}^*}{h_\omega} \right) (\hat{g}_{it} - g_{it} - \hat{g}_{jt} + g_{jt}) \right. \\ & - \left. \kappa_g \left(\frac{g_{ijt}^*}{h_\omega} \right) \kappa_v^{(1)} \left(\frac{v_{ijt2}^*}{h_\omega} \right) (\mathbf{q}_{it} - \mathbf{q}_{jt})(\hat{\boldsymbol{\pi}}_t - \boldsymbol{\pi}_t) \right\} d_{it} \cdot d_{jt} \cdot \tau_{it} \cdot \tau_{jt}, \end{aligned} \tag{8}$$

where $\kappa_g^{(1)}(\cdot)$ and $\kappa_v^{(1)}(\cdot)$ are vectors of first derivatives of functions $\kappa_g(\cdot)$ and $\kappa_v(\cdot)$, respectively, $\mathbf{q}_{it} = (1, \mathbf{z}_{i1}, \ldots, \mathbf{z}_{it})$, $\boldsymbol{\pi}_t = (\eta_{2t}, \boldsymbol{\xi}_{21}, \ldots, \boldsymbol{\delta}_t + \boldsymbol{\xi}_{2t}, \ldots, \boldsymbol{\xi}_{2T})'$, and $\hat{\boldsymbol{\pi}}_t$ is the first-step estimator of $\boldsymbol{\pi}_t$.

Similar to Blundell and Powell (2004), if the first four moments of $\mathbf{r}_{it}$ and s_{it} are finite, and $\kappa_g(\cdot)$, $\kappa_v(\cdot)$, τ_{it}, τ_{jt} are bounded, then $\hat{\mathbf{S}}_0^t = \boldsymbol{\Sigma}_0^t + o_p(1)$, $t = 1, ..., T$, when $h_\omega \to 0$, $h_\omega^2 N \to \infty$.

To show that $\mathbf{S}_1^t$ converges in probability to zero, $t = 1, \ldots, T$, assume that functions $\kappa_g(\cdot)$, $\kappa_v(\cdot)$, $\kappa_g^{(1)}(\cdot)$, $\kappa_v^{(1)}(\cdot)$ are uniformly bounded, and the first two moments of $\mathbf{q}_{it}$ exist. For the first-step Powell's censored least absolute deviations estimator (Powell, 1984) or symmetrically trimmed censored least squares estimator (Powell, 1986), assume that the appropriate regularity conditions hold, so that $\hat{\boldsymbol{\pi}}_t$ is $\sqrt{N}$-consistent for all t. Moreover, assume that regularity conditions provided in Ahn and Powell (1993) are satisfied. These include smoothness assumptions for conditional expectation and density functions conditional on $g_{it} = g_{jt}$ and $v_{it2} = v_{jt2}$, as well as the restrictions on the speed with which h_g and h_ω converge to zero as $N \to \infty$, where both depend on the dimensionality of the continuous component of w_{it}. There is also a requirement that higher-order (bias-reducing) kernels are used at both steps. The second-step kernels, κ_g and κ_v, are assumed to be fourth-order kernels with the first three moments being equal to zero. For the first-step kernel, K, the number of vanishing moments depends on the number of continuous variables in w_{it}. If these assumptions hold, the biases resulting from the nonparametric estimation of g_{it} and ω_{ijt} are of the order smaller than $\sqrt{N}$.

From above, it follows that under the specified conditions,

$$\hat{\mathbf{S}} \equiv \sum_{t=1}^{T} \hat{\mathbf{S}}^t = \sum_{t=1}^{T} \mathbf{\Sigma}_0^t + o_p(1) \equiv \mathbf{\Sigma}_0 + o_p(1). \tag{9}$$

Moreover, using the law of iterated expectations:

$$\begin{aligned} \mathbf{\Sigma}_0^t &\equiv \mathrm{E}[f_{it} \cdot d_{it} \cdot d_{jt} \cdot \tau_{it} \cdot \tau_{jt} \cdot (\mathbf{w}_{it} - \mathbf{w}_{jt})'(\mathbf{w}_{it} - \mathbf{w}_{jt}) | g_{it} = g, v_{it2} = v] \\ &= \mathrm{E}\left\{2 f_{it} \cdot (\varrho_{it}\mu_{ww,it} - \mu'_{w,it}\mu_{w,it})\right\}, \quad t = 1, \ldots, T, \end{aligned} \tag{10}$$

$$\begin{aligned} \varrho_{it} &\equiv \mathrm{E}[d_{it} \cdot \tau_{it} | g_{it} = g, v_{it2} = v], \\ \mu_{w,it} &\equiv \mathrm{E}[d_{it} \cdot \tau_{it} \cdot \mathbf{w}_{it} | g_{it} = g, v_{it2} = v], \\ \mu_{ww,it} &\equiv \mathrm{E}[d_{it} \cdot \tau_{it} \cdot \mathbf{w}'_{it}\mathbf{w}_{it} | g_{it} = g, v_{it2} = v]. \end{aligned} \tag{11}$$

Furthermore, $\mathbf{\Sigma}_0 \boldsymbol{\theta} = 0$ because

$$\begin{aligned} \sum_{t=1}^{T} [\varrho_{it}\mu_{ww,it} - \mu'_{w,it}\mu_{w,it}]\boldsymbol{\theta} &= \sum_{t=1}^{T} \left[\varrho_{it}\mathrm{E}(\mathbf{w}'_{it}\mathbf{w}_{it}\boldsymbol{\theta} | g_{it}, v_{it2}) - \mu'_{w,it}\mathrm{E}(\mathbf{w}_{it}\boldsymbol{\theta} | g_{it}, v_{it2})\right] \\ &= \sum_{t=1}^{T} (\varrho_{it}\mu'_{w,it} g_{it} - \varrho_{it}\mu'_{w,it} g_{it}) = 0, \end{aligned} \tag{12}$$

where we use the fact that $\mathbf{w}_{it}\boldsymbol{\theta} = g_{it}$, $t = 1, \ldots, T$.

Finally, the identification condition has to hold. Regarding the first-step estimation, necessary identification conditions for the censored least absolute deviations estimator and symmetrically trimmed least squares estimator are provided in Powell (1984) and Powell (1986), respectively. The second part of the identification condition is that in the population, $\boldsymbol{\theta}$ is a unique nontrivial solution to $\mathbf{\Sigma}_0 \boldsymbol{\theta} = 0$ after the normalization $\boldsymbol{\theta} = (1, \alpha')'$ is imposed. Specifically, assume that matrix $\mathbf{\Sigma}_0^{22}$, which is the lower-right $(M + L - 1) \times (M + L - 1)$ sub-matrix of matrix $\mathbf{\Sigma}_0$, has full rank. This completes the consistency argument.

In order to establish $\sqrt{N}$-asymptotic normality, we first use the second order mean value expansion to write

$$\hat{\mathbf{S}} = \mathbf{S}_0 + \mathbf{S}_1 + \mathbf{S}_2 \equiv \sum_{t=1}^{T} \mathbf{S}_0^t + \sum_{t=1}^{T} \mathbf{S}_1^t + \sum_{t=1}^{T} \mathbf{S}_2^t, \tag{13}$$

where

$$\mathbf{S}_l^t \equiv \begin{pmatrix} n \\ 2 \end{pmatrix}^{-1} \sum\nolimits_{i<j} \omega_{ijt}^l (\mathbf{w}_{it} - \mathbf{w}_{jt})'(\mathbf{w}_{it} - \mathbf{w}_{jt}),\ l = 0, 1, \tag{14}$$

$$\omega_{ijt}^0 \equiv \frac{1}{h_\omega^2} \kappa_g \left(\frac{g_{it} - g_{jt}}{h_\omega} \right) \kappa_v \left(\frac{v_{it2} - v_{jt2}}{h_\omega} \right) d_{it} \cdot d_{jt} \cdot \tau_{it} \cdot \tau_{jt}, \tag{15}$$

$$\begin{aligned} \omega_{ijt}^1 \equiv\ & \frac{1}{h_\omega^3} \Bigg\{ \kappa_g^{(1)} \left(\frac{g_{it} - g_{jt}}{h_\omega} \right) \kappa_v \left(\frac{v_{it2} - v_{jt2}}{h_\omega} \right) (\hat{g}_{it} - g_{it} - \hat{g}_{jt} + g_{jt}) \\ & - \kappa_g \left(\frac{g_{it} - g_{jt}}{h_\omega} \right) \kappa_v^{(1)} \left(\frac{v_{it2} - v_{jt2}}{h_\omega} \right) (\mathbf{q}_{it} - \mathbf{q}_{jt})(\hat{\boldsymbol{\pi}}_t - \boldsymbol{\pi}_t) \Bigg\} d_{it} \cdot d_{jt} \cdot \tau_{it} \cdot \tau_{jt}, \end{aligned} \tag{16}$$

$$\begin{aligned} \omega_{ijt}^2 \equiv\ & \frac{1}{2h_\omega^4} \Bigg\{ \kappa_g^{(2)} \left(\frac{g_{ijt}^*}{h_\omega} \right) \kappa_v \left(\frac{v_{ijt2}^*}{h_\omega} \right) (\hat{g}_{it} - g_{it} - \hat{g}_{jt} + g_{jt})^2 \\ & - 2\kappa_g^{(1)} \left(\frac{g_{ijt}^*}{h_\omega} \right) \kappa_v^{(1)} \left(\frac{v_{ijt2}^*}{h_\omega} \right) \cdot (\mathbf{q}_{it} - \mathbf{q}_{jt})(\hat{\boldsymbol{\pi}}_t - \boldsymbol{\pi}_t)(\hat{g}_{it} - g_{it} - \hat{g}_{jt} + g_{jt}) \\ & + \kappa_g \left(\frac{g_{ijt}^*}{h_\omega} \right) \kappa_v^{(2)} \left(\frac{v_{ijt2}^*}{h_\omega} \right) (\mathbf{q}_{it} - \mathbf{q}_{jt})(\hat{\boldsymbol{\pi}}_t - \boldsymbol{\pi}_t)(\hat{\boldsymbol{\pi}}_t - \boldsymbol{\pi}_t)'(\mathbf{q}_{it} - \mathbf{q}_{jt})' \Bigg\} d_{it} \cdot d_{jt} \cdot \tau_{it} \cdot \tau_{jt} \end{aligned} \tag{17}$$

Under assumptions stated in Ahn and Powell (1993), using $\sqrt{N}$-consistency of the first-step estimator $\hat{\boldsymbol{\pi}}$, and following the same argument as in Blundell and Powell (2004), it should be the case that

$$\sqrt{N}\mathbf{S}_0\boldsymbol{\theta} = o_p(1), \quad \sqrt{N}\mathbf{S}_2\boldsymbol{\theta} = o_p(1). \tag{18}$$

Furthermore, when the selection equation is estimated using either Powell's censored least absolute deviations estimator or symmetrically trimmed censored least squares estimator, $\hat{\boldsymbol{\pi}}$ satisfies

$$\sqrt{N}(\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{m}_i + o_p(1),$$

where $\mathrm{E}(\mathbf{m}_i) = 0$, and $\mathrm{E}(\mathbf{m}_i\mathbf{m}_i')$ exists and is nonsingular.

Then, we can show that

$$\sqrt{N}\hat{\mathbf{S}}\boldsymbol{\theta} = \sqrt{N}\mathbf{S}_1\boldsymbol{\theta} + o_p(1) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{e}_{i1} + \mathbf{e}_{i2}) + o_p(1), \tag{19}$$

where

$$
\begin{aligned}
\mathbf{e}_{i1} &\equiv \sum_{t=1}^{T} 2 f_{it} \varrho_{it} (\varrho_{it} \mathbf{w}_{it} - \mu_{w,it})' \cdot \frac{\partial \psi(g_{it}, v_{it2})}{\partial\ g_{it}} \cdot [y_{it} - g(\mathbf{r}_{it})], \\
\mathbf{e}_{i2} &\equiv -\mathbf{F}\mathbf{m}_i(\boldsymbol{\pi}), \\
\mathbf{F} &\equiv \mathrm{E}\left[\sum_{t=1}^{T} 2 f_{it} \varrho_{it} (\varrho_{it} \mathbf{w}_{it} - \mu_{w,it})' \cdot \frac{\partial \psi(g_{it}, v_{it2})}{\partial\ v_{it2}} \cdot \mathbf{q}_{it}\right].
\end{aligned} \tag{20}
$$

If the censored least absolute deviations estimator (Powell, 1984) is used as the estimator of $\boldsymbol{\pi}$, and the first-step estimation is performed separately for each t, then

$$
\begin{aligned}
\mathbf{m}_i(\boldsymbol{\pi}) &= \begin{pmatrix} \mathbf{m}_{i1}(\boldsymbol{\pi}_1) \\ \dots \\ \mathbf{m}_{iT}(\boldsymbol{\pi}_T) \end{pmatrix}, \\
\mathbf{m}_{it}(\boldsymbol{\pi}_t) &= [\mathrm{f}_t(0) \cdot \mathbf{J}_t]^{-1} \cdot 1[\mathbf{q}_{it}\boldsymbol{\pi}_t > 0] \cdot \mathbf{q}'_{it} \left(\frac{1}{2} - 1[v_{it2} > 0]\right), \\
\mathbf{J}_t &\equiv \mathrm{E}\left(1[\mathbf{q}_{it}\boldsymbol{\pi}_t > 0] \cdot \mathbf{q}'_{it}\mathbf{q}_{it}\right), \quad t = 1, \dots, T,
\end{aligned} \tag{21}
$$

where $\mathrm{f}_t(\cdot)$ is the density function of error v_{it2} in period t.

If $\boldsymbol{\pi}_t$, $t = 1, \dots, T$, is estimated using the symmetrically trimmed censored least squares estimator (Powell, 1986), then

$$
\begin{aligned}
\mathbf{m}_{it}(\boldsymbol{\pi}_t) &= \mathbf{C}_t^{-1} \cdot 1[\mathbf{q}_{it}\boldsymbol{\pi}_t > 0] \cdot \mathbf{q}'_{it} \cdot \left(\min\{s_{it}, 2\mathbf{q}_{it}\boldsymbol{\pi}_t\} - \mathbf{q}_{it}\boldsymbol{\pi}_t\right), \\
\mathbf{C}_t &\equiv \mathrm{E}\left\{1[-\mathbf{q}_{it}\boldsymbol{\pi}_t < v_{it2} < \mathbf{q}_{it}\boldsymbol{\pi}_t] \cdot \mathbf{q}'_{it}\mathbf{q}_{it}\right\}, \quad t = 1, \dots, T.
\end{aligned} \tag{22}
$$

From (12) and (19) it follows that

$$\sqrt{N}\boldsymbol{\theta}'\hat{\mathbf{S}}\boldsymbol{\theta} = o_p(1), \tag{23}$$

so that for the subvector $\hat{\alpha}$ of $\hat{\boldsymbol{\theta}} = (1, \hat{\alpha'})'$, we obtain

$$\sqrt{N}(\hat{\alpha} - \alpha) \xrightarrow{d} Normal(0, \boldsymbol{\Sigma}_{22}^{-1}\mathbf{V}_{22}\boldsymbol{\Sigma}_{22}^{-1}), \tag{24}$$

where $\boldsymbol{\Sigma}_{22}$ is the lower $(M + L - 1) \times (M + L - 1)$ diagonal submatrix of $\boldsymbol{\Sigma}_0$, and $\mathbf{V}_{22}$ is the lower $(M + L - 1) \times (M + L - 1)$ diagonal submatrix of $\mathbf{V}$,

$$\mathbf{V} \equiv \mathrm{Var}(\mathbf{e}_{i1} + \mathbf{e}_{i2}) = \mathrm{E}[(\mathbf{e}_{i1} + \mathbf{e}_{i2})(\mathbf{e}_{i1} + \mathbf{e}_{i2})']. \tag{25}$$

Note that this is a robust form of the variance that accounts for serial dependence in the errors.

To obtain a consistent estimator of $\text{Avar}[\sqrt{N}(\hat{\alpha}-\alpha)]$, first note that $\hat{S}$ is consistent for Σ_0. Furthermore, using the argument similar to the one in Ahn and Powell (1993), a consistent estimator of V would be

$$\hat{V} \equiv \frac{1}{N}\sum_{i=1}^{N}[(\hat{\mathbf{e}}_{i1}+\hat{\mathbf{e}}_{i2})(\hat{\mathbf{e}}_{i1}+\hat{\mathbf{e}}_{i2})'], \tag{26}$$

where

$$\begin{aligned} \hat{\mathbf{e}}_{i1} &\equiv \sum_{t=1}^{T}\frac{1}{N-1}\sum_{j=1}^{N}\left[\frac{2}{h_w^2}K_g'\left(\frac{\hat{g}_{it}-\hat{g}_{jt}}{h_w}\right)K_v\left(\frac{\hat{v}_{it2}-\hat{v}_{jt2}}{h_w}\right)d_{it}d_{jt}\tau_{it}\tau_{jt}\hat{\delta}_{jt}\right](y_{it}-\hat{g}_{it}), \\ \hat{\delta}_{jt} &\equiv \frac{\sum_{l=1}^{N}K\left(\frac{\hat{\mathbf{r}}_{jt}-\hat{\mathbf{r}}_{lt}}{h_g}\right)(\mathbf{w}_{jt}-\mathbf{w}_{lt})'}{\sum_{l=1}^{N}K\left(\frac{\hat{\mathbf{r}}_{jt}-\hat{\mathbf{r}}_{lt}}{h_g}\right)}, \end{aligned} \tag{27}$$

and

$$\begin{aligned} \hat{\mathbf{e}}_{i2} &\equiv -\hat{\mathbf{F}}\mathbf{m}_i(\hat{\boldsymbol{\pi}}), \\ \hat{\mathbf{F}} &\equiv \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{1}{N-1}\sum_{j=1}^{N}\left[\frac{2}{h_w^2}K_g\left(\frac{\hat{g}_{it}-\hat{g}_{jt}}{h_w}\right)K_v'\left(\frac{\hat{v}_{it2}-\hat{v}_{jt2}}{h_w}\right)d_{it}d_{jt}\tau_{it}\tau_{jt}\hat{\delta}_{jt}\right]\mathbf{q}_{it}, \end{aligned} \tag{28}$$

for $\mathbf{m}_i(\hat{\boldsymbol{\pi}}) = [\mathbf{m}_{i1}(\hat{\boldsymbol{\pi}}_1),\ldots,\mathbf{m}_{iT}(\hat{\boldsymbol{\pi}}_T)]'$, and $\mathbf{m}_{it}(\hat{\boldsymbol{\pi}}_t)$ defined as in either (21) or (22), but evaluated at $\hat{\boldsymbol{\pi}}_t$.

References

Ahn, H. and Powell, J.L., 1993, Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism. *Journal of Econometrics* 58, 3-29.

Blundell, R.W. and Powell, J.L. , 2004, Endogeneity in Semiparametric Binary Response Models. *Review of Economic Studies* 71, 655-679.

Powell, J.L., 1984, Least Absolute Deviations Estimation for the Censored Regression Model. *Journal of Econometrics* 25, 303-325.

Powell, J.L., 1986, Symmetrically Trimmed Least Squares Estimation for Tobit Models. *Econometrica* 54, 1435-1460.

Table 1: Simulation results for $\hat{\beta}_2/\hat{\beta}_1$ ($\beta_2/\beta_1 = 0.6$), $u_{it1} \sim \chi^2_3$

		Probit (1)	Probit, time means (2)	s_{it} censored, 2-step MLE (3)	s_{it} binary, full MLE (4)	s_{it} censored, Semiparametric (5)
				$\sigma^2_a = 0$, $\xi_1 = 0$, $\rho = 0$		
N=500	Bias	0.0006	0.0017	0.0006	0.0052	0.0051
	RMSE	0.0482	0.0584	0.0612	0.0594	0.0698
	Average se	0.0477	0.0583	0.0613	0.0594	0.0693
	Bootstrap se					0.0815
				$\sigma^2_a = 0.3$, $\xi_1 = -0.3$, $\rho = 0$		
N=500	Bias	-0.0630	0.0000	-0.0008	0.0016	0.0098
	RMSE	0.0912	0.0654	0.0698	0.0672	0.0812
	Average se	0.0641	0.0673	0.0711	0.0684	0.0795
	Bootstrap se					0.0945
				$\sigma^2_a = 0.3$, $\xi_1 = -0.3$, $\rho = 0.5$		
N=500	Bias	-0.1237	-0.0459	0.0011	0.0042	-0.0048
	RMSE	0.1402	0.0816	0.0698	0.0677	0.0779
	Average se	0.0680	0.0690	0.0712	0.0687	0.0764
	Bootstrap se					0.0988
				$\sigma^2_a = 0$, $\xi_1 = 0$, $\rho = 0$		
N=1000	Bias	0.0009	0.0021	0.0007	0.0052	0.0057
	RMSE	0.0350	0.0426	0.0446	0.0438	0.0471
	Average se	0.0338	0.0413	0.0434	0.0420	0.0509
	Bootstrap se					0.0539
				$\sigma^2_a = 0.3$, $\xi_1 = -0.3$, $\rho = 0$		
N=1000	Bias	-0.0617	0.0006	-0.0002	0.0020	0.0070
	RMSE	0.0775	0.0495	0.0517	0.0506	0.0557
	Average se	0.0455	0.0476	0.0502	0.0484	0.0476
	Bootstrap se					0.0623
				$\sigma^2_a = 0.3$, $\xi_1 = -0.3$, $\rho = 0.5$		
N=1000	Bias	-0.1213	-0.0455	0.0004	0.0037	-0.0026
	RMSE	0.1310	0.0672	0.0510	0.0501	0.0529
	Average se	0.0482	0.0489	0.0503	0.0485	0.0573
	Bootstrap se					0.0643
				$\sigma^2_a = 0$, $\xi_1 = 0$, $\rho = 0$		
N=2500	Bias	0.0017	0.0018	-0.00002	0.0049	0.0033
	RMSE	0.0218	0.0262	0.0274	0.0270	0.0294
	Average se	0.0214	0.0261	0.0274	0.0265	0.0469
				$\sigma^2_a = 0.3$, $\xi_1 = -0.3$, $\rho = 0$		
N=2500	Bias	-0.0605	0.0023	0.0008	0.0033	0.0044
	RMSE	0.0672	0.0303	0.0326	0.0310	0.0357
	Average se	0.0288	0.0301	0.0318	0.0306	0.0275
				$\sigma^2_a = 0.3$, $\xi_1 = -0.3$, $\rho = 0.5$		
N=2500	Bias	-0.1215	-0.0458	0.0002	0.0032	-0.0047
	RMSE	0.1255	0.0550	0.0309	0.0301	0.0327
	Average se	0.0305	0.0309	0.0317	0.0306	0.0476

Table 2: Descriptive Statistics for NLSY79 data

Variable	Mean
Age (years)	31.07
	(2.63)
Education (years)	13.51
	(2.35)
AFQT score	54.05
	(26.16)
Married (%)	69.39
Number of observations	8,340

Sample standard deviations are in parentheses below the sample means