Supplement to
"Yet Another Look at the Omitted Variable Bias"

Masayuki Hirukawa* Ryukoku University

Irina Murtazashvili† Drexel University

Artem Prokhorov‡ University of Sydney & CEBA & CIREQ

December 6, 2022

1 Alternative Estimators to PILS

This section discusses the definitions and statistical properties of estimators alternative to PILS. The estimators include IV, PARA and three matching-based estimators.

1.1 IV Estimator

One may attempt to complete the estimation using only $\mathcal{S}_1$, for example, in light of a cost associated with searching for another data source containing $\mathcal{S}_2$. In this case, X_2 in

$$Y = \beta_0 + X_1'\beta_1 + X_2'\beta_2 + X_{3I}'\beta_3 + u, \tag{1}$$

is treated as the vector of omitted variables. Accordingly, regression (1) reduces to $Y = X_S'\beta_S + v$, where $X_S := (1, X_1', X_{3I}')'$, $\beta_S := (1, \beta_1', \beta_3')'$ and $v := u + X_2'\beta_2$. The subscript "S" stands for a short regression.

OLS for the short regression is inconsistent if (X_1, X_{3I}) and X_2 are correlated. Suppose that $\mathcal{S}_1$ additionally includes a vector of instruments $Z \in \mathbb{R}^{d_Z}$ that correlate

*e-mail: hirukawa@econ.ryukoku.ac.jp.
†e-mail: im99@drexel.edu.
‡e-mail: artem.prokhorov@sydney.edu.au.

with (X_1, X_{3I}) but not with X_2. Note that X_{3E} is not used as instruments unless it is part of Z. Then, it is possible to estimate β_S consistently on the basis of the moment restriction

$$E\left(Z_S v\right) := E\left\{\begin{bmatrix} 1 \\ Z \end{bmatrix}\left(u + X_2' \beta_2\right)\right\} = 0_{(d_Z+1)\times 1}.$$

Assuming $d_Z \geq d_1 + d_{3I}$, consistency of this estimator will be based on strong assumptions such as $Z \perp\!\!\!\perp X_2$ and $E\left(X_2\right)' \beta_2 = 0$. The benefit is a simple estimator which, in the just-identified case ($d_Z = d_1 + d_{3I}$), can be written as

$$\hat{\beta}_{IV} := S^{-1}_{n,Z_S,X_S} S_{n,Z_S,Y} := \left(\frac{1}{n}\sum_{i=1}^{n} Z_{S,i} X'_{S,i}\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_{S,i} Y_i\right).$$

Under suitable regularity conditions, $\hat{\beta}_{IV} \xrightarrow{p} \beta_S$ and

$$\sqrt{n}\left(\hat{\beta}_{IV} - \beta_S\right) \xrightarrow{d} N\left(0_{(d_Z+1)\times 1}, \Phi^{-1}_{Z_S,X_S} \Upsilon \Phi'^{-1}_{Z_S,X_S}\right),$$

where $\Upsilon := E\left(Z_S Z_S' v^2\right)$.

1.2 PARA Estimator

Instead of imputing a nonparametric estimate of $g_2\left(X_3\right)$ in place of the missing regressor X_2 in regression (1), Fang, Keane, and Silverman (2008) propose to employ a parametric, linear predictor of X_2. Since Fang, Keane, and Silverman (2008) do not document statistical properties of PARA, we explore them below.

PARA is a two-step two-sample estimator. Consider the linear projection of X_2 on X_3, i.e., $\text{proj}\left(X_2 | X_3\right) := A\mathcal{X}_3$, where $\mathcal{X}_3 := (1, X_3')'$ and A is some $d_2 \times (d_3+1)$ coefficient matrix. In the first step, we estimate A using $\mathcal{S}_2$ by means of equation-by-equation OLS. As a result, n linear predictors $\left\{\tilde{X}_{2i}\right\}_{i=1}^{n} = \left\{\tilde{A}\mathcal{X}_{3i}\right\}_{i=1}^{n}$ can be obtained, where $\tilde{A}$ is the OLS estimate of A. Then, in the second step, we run OLS for the regression of Y on $\tilde{X} := \left(1, X_1', \tilde{X}_2', X_{3I}'\right)'$ using $\mathcal{S}_1$ to estimate β. The PARA estimator is then defined as

$$\hat{\beta}_{PARA} := S^{-1}_{n,\tilde{X},\tilde{X}} S_{n,\tilde{X},Y} = \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i Y_i.$$

The first proposition refers to the condition for consistency of PARA. In what follows, denote $\mathcal{X}^A := \left(1, X_1', (A\mathcal{X}_3)', X_{3I}'\right)'$.

Proposition S1. *Suppose that* $d_{3E} > 0$ *and that* X_2 *is relevant in regression (1), i.e.,* $\beta_2 \neq 0$*. Then, under Assumptions 1, 2 and 3(i) of the paper,* $\hat{\beta}_{PARA} \xrightarrow{p} \beta$ *as* $n, m \to \infty$ *if and only if* $E\left(X_1 e'\right) = 0_{d_1 \times d_2}$*, where* $e := X_2 - A\mathcal{X}_3$ *is the projection error.*

Proof. By construction, e is orthogonal to $\mathcal{X}_3$, i.e., $E\left(\mathcal{X}_3 e'\right) = 0$. We note that this is true even if $\operatorname{proj}\left(X_2 | X_3\right) \neq g_2\left(X_3\right) = E\left(X_2 | X_3\right)$, which would be the case if the conditional expectation function is nonlinear in X_3. A simple manipulation of regression (1) yields $Y = \beta_0 + X_1'\beta_1 + \tilde{X}_2'\beta_2 + X_{3I}'\beta_3 + \varepsilon$, where $\varepsilon := u + e'\beta_2 - \mathcal{X}_3'\left(\tilde{A} - A\right)'\beta_2$ is the composite regression error. Also observe that $\tilde{A} = A + O_p\left(m^{-1/2}\right) = A + o_p(1)$. Then, $E\left(\tilde{X}_2\varepsilon\right) = E\left(A\mathcal{X}_3\varepsilon\right) + E\left\{\left(\tilde{A} - A\right)\mathcal{X}_3\varepsilon\right\} = E\left(A\mathcal{X}_3\varepsilon\right) + o(1)$, and PARA is consistent if and only if $E\left(\mathcal{X}_i^A\varepsilon_i\right) = 0_{(d+1)\times 1}$. By Assumption 3(i) of the paper, $\tilde{A} \perp\!\!\!\perp \mathcal{S}_1$ and $E\left(\tilde{A}\right) = A$, we have $E\left(\varepsilon_i\right) = 0$, $E\left(A\mathcal{X}_{3i}\varepsilon_i\right) = 0_{d_2\times 1}$, and $E\left(X_{3Ii}\varepsilon_i\right) = 0_{d_{3I}\times 1}$. However, $E\left(X_{1i}\varepsilon_i\right) = E\left(X_{1i}e_i'\right)\beta_2 \neq 0_{d_1\times 1}$ unless $E\left(X_1 e'\right) = 0_{d_1\times d_2}$. ■

Proposition S1 suggests that the distinction between X_1 and X_3 becomes important. PARA is consistent if X_1 is empty. If X_1 is non-empty, then consistency of PARA is not guaranteed and depends on whether or not $E\left(X_1 e'\right) = 0_{d_1\times d_2}$. We may be tempted to test the null of $E\left(X_1 e'\right) = 0_{d_1\times d_2}$ before running PARA. However, since X_1 and X_2 belong to two distinct samples, this task may not be feasible.

In the simulation study in Section 2 of this Supplement, we consider the cases in which PARA is consistent although X_1 is non-empty in regression (1). The proposition below justifies such cases.

Proposition S2. (Consistency of PARA in Simulations) *Suppose that X_1 is non-empty in regression (1). If the function $h(\cdot)$ used in Simulation Studies I and III of this Supplement is odd, then PARA is consistent.*

Proof. We focus only on the design for Simulation Study I to save space and adopt the notations used there. Let $\operatorname{proj}(X_2 | X_3) = A\mathcal{X}_3 = A_0 + A_1 X_{3I_C} + A_2 X_{3E_C} + A_3 X_{3E_D}$. Then, since X_{3I_C}, X_{3E_C}, X_{3E_D}, and η_2 are zero-mean mutually independent random variables, the reduced form of X_2 implies that all the following statements are true:

1. $A_0 = 0$, $A_2 = -(3/4)(H_2 + 1)$, $A_3 = 2$;

2. A_1 satisfies $A_1 X_{3I_C} = \operatorname{proj}\{h(X_{3I_C}) | X_{3I_C}\}$; and

3. The projection error $e = X_2 - A\mathcal{X}_3 = h(X_{3I_C}) - A_1 X_{3I_C} + \eta_2$.

By Proposition S1 above, PARA is consistent if and only if $E(X_1 e) = 0$. However, it follows from the reduced form of X_1 and distributional assumptions on X_{3I_C}, X_{3E_C}, X_{3E_D}, and η_1 that $E(X_1 e) = H_2 = E\left\{X_{3I_C}^2 h(X_{3I_C})\right\}$ holds. The right-hand side is zero if and only if $h(\cdot)$ is odd, which establishes the proposition. ■

We conclude this subsection by documenting the limiting distribution of PARA. The covariance estimate if $n/m \to \kappa$ is applied in Sections 3 and 4 of the paper, as well as Sections 2 and 3 of this Supplement.

Proposition S3. *If the same regularity conditions as in Proposition S1 hold,* $E\left(X_1 e'\right) = 0_{d_1 \times d_2}$, $\Phi_{\mathcal{X}^A,\mathcal{X}^A} = E\left(\mathcal{X}^A \mathcal{X}^{A\prime}\right) > 0$, *and* $\Phi_{\mathcal{X}_3,\mathcal{X}_3} = E\left(\mathcal{X}_3 \mathcal{X}_3'\right) > 0$, *then, as* $n, m \to \infty$,

$$\begin{cases} \sqrt{n}\left(\hat{\beta}_{PARA} - \beta\right) \xrightarrow{d} N\left(0_{(d+1)\times 1}, \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A} \Psi \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\right) & \text{if } n/m \to \kappa \in (0,\infty) \\ \sqrt{n}\left(\hat{\beta}_{PARA} - \beta\right) \xrightarrow{d} N\left(0_{(d+1)\times 1}, \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A} \Psi_1 \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\right) & \text{if } n/m \to 0 \\ \sqrt{m}\left(\hat{\beta}_{PARA} - \beta\right) \xrightarrow{d} N\left(0_{(d+1)\times 1}, \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A} \Psi_2 \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\right) & \text{if } n/m \to \infty \end{cases},$$

where $\Psi := \Psi_1 + \kappa \Psi_2$, $\Psi_1 := E\left(\mathcal{X}^A \mathcal{X}^{A\prime} v^2\right)$, $v := u + e'\beta_2$, *and*

$$\Psi_2 := \Phi_{\mathcal{X}^A,\mathcal{X}_3} \Phi^{-1}_{\mathcal{X}_3,\mathcal{X}_3} E\left(\mathcal{X}_3 \mathcal{X}_3' \beta_2' e e' \beta_2\right) \Phi^{-1}_{\mathcal{X}_3,\mathcal{X}_3} \Phi'_{\mathcal{X}^A,\mathcal{X}_3}.$$

In addition, asymptotic variances Ψ, Ψ_1 and Ψ_2 *can be consistently estimated by* $\hat{\Psi} := \hat{\Psi}_1 + (n/m)\hat{\Psi}_2$, *where* $\hat{\Psi}_1 := (1/n)\sum_{i=1}^n \tilde{X}_i \tilde{X}_i' \hat{v}_i^2$, $\hat{v}_i := Y_i - \tilde{X}_i' \hat{\beta}_{PARA}$ *is the PARA residual,*

$$\hat{\Psi}_2 := S_{n,\tilde{X},\mathcal{X}_3} S^{-1}_{m,\mathcal{X}_3,\mathcal{X}_3} \left(\frac{1}{m}\sum_{j=1}^m \mathcal{X}_{3j}\mathcal{X}_{3j}' \hat{\beta}'_{PARA,2} \tilde{e}_j \tilde{e}_j' \hat{\beta}_{PARA,2}\right) S^{-1}_{m,\mathcal{X}_3,\mathcal{X}_3} S'_{n,\tilde{X},\mathcal{X}_3},$$

$\tilde{e}_j := X_{2j} - \tilde{A} X_{3j}$ *is the first-step OLS residual, and* $\hat{\beta}_{PARA,2}$ *is the PARA estimate of* β_2.

Proof. The proof starts from the first-step equation-by-equation OLS estimation. The estimate of A using $\mathcal{S}_2$ is given by

$$\tilde{A} = S_{m,X_2,\mathcal{X}_3} S_{m,\mathcal{X}_3,\mathcal{X}_3} = \frac{1}{m}\sum_{j=1}^m X_{2j}\mathcal{X}_{3j}' \left(\frac{1}{m}\sum_{j=1}^m \mathcal{X}_{3j}\mathcal{X}_{3j}'\right)^{-1}.$$

Substituting $X_{2j} = A\mathcal{X}_{3j} + e_j$ into this yields

$$\tilde{A} = A + \frac{1}{m}\sum_{j=1}^m e_j \mathcal{X}_{3j}' \left(\frac{1}{m}\sum_{j=1}^m \mathcal{X}_{3j}\mathcal{X}_{3j}'\right)^{-1}. \tag{2}$$

Now we proceed to the second step. Denoting $\tilde{X}_2 = \tilde{A}\mathcal{X}_3$ and $\tilde{X} = \left(1, X_1', \tilde{X}_2', X_{3I}'\right)'$, we can rewrite the regression model $Y = \beta_0 + X_1'\beta_1 + X_2'\beta_2 + X_{3I}'\beta_3 + u$ as $Y = \tilde{X}'\beta + \varepsilon$, where

$$\varepsilon = v - \mathcal{X}_3'\left(\tilde{A} - A\right)'\beta_2 = (u + e'\beta_2) - \mathcal{X}_3'\left(\tilde{A} - A\right)'\beta_2.$$

It follows that

$$\begin{aligned}
\hat{\beta}_{PARA} &= S^{-1}_{n,\tilde{X},\tilde{X}} S_{n,\tilde{X},Y} \\
&= \beta + S^{-1}_{n,\tilde{X},\tilde{X}} S_{n,\tilde{X},v} - S^{-1}_{n,\tilde{X},\tilde{X}} S_{n,\tilde{X},\mathcal{X}_3'\left(\tilde{A}-A\right)'\beta_2} \\
&= \beta + \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \tilde{X}_i' \right)^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i v_i - \frac{1}{n} \sum_{i=1}^{n} \tilde{X}_i \mathcal{X}_{3i}' \left(\tilde{A} - A \right)' \beta_2 \right\}. \qquad (3)
\end{aligned}$$

Because $\tilde{A} = A + O_p\left(m^{-1/2}\right) = A + o_p(1)$ by (2),

$$S_{n,\tilde{X},\tilde{X}} = S_{n,\mathcal{X}^A,\mathcal{X}^A} + o_p(1) = \frac{1}{n} \sum_{i=1}^{n} \mathcal{X}_i^A \mathcal{X}_i^{A\prime} + o_p(1) \xrightarrow{p} \Phi_{\mathcal{X}^A,\mathcal{X}^A}, \qquad (4)$$

where $\mathcal{X}^A = \left(1, X_1', (A\mathcal{X}_3)', X_{3I}'\right)'$. Also, by CLT,

$$\begin{aligned}
\sqrt{n} S_{n,\tilde{X},v} &= \sqrt{n} S_{n,\mathcal{X}^A,v} + \sqrt{n} S_{n,\tilde{X}-\mathcal{X}^A,v} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_i^A v_i + O\left\{ \left(\tilde{A} - A \right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_{3i} v_i \right\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_i^A v_i + o_p(1) \\
&\xrightarrow{d} N\left(0_{(d+1)\times 1}, \Psi_1\right), \qquad (5)
\end{aligned}$$

where $\Psi_1 = E\left(\mathcal{X}^A \mathcal{X}^{A\prime} v^2\right)$.

Moreover,

$$\begin{aligned}
&\sqrt{n} S_{n,\tilde{X},\mathcal{X}_3'\left(\tilde{A}-A\right)'\beta_2} \\
&= \sqrt{n} S_{n,\mathcal{X}^A,\mathcal{X}_3'\left(\tilde{A}-A\right)'\beta_2} + \sqrt{n} S_{n,\tilde{X}-\mathcal{X}^A,\mathcal{X}_3'\left(\tilde{A}-A\right)'\beta_2} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_i^A \mathcal{X}_{3i}' \left(\tilde{A} - A \right)' \beta_2 + O\left\{ \left(\tilde{A} - A \right) \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_{3i} \mathcal{X}_{3i}' \left(\tilde{A} - A \right)' \beta_2 \right\} \\
&= \sqrt{\frac{n}{m}} \left\{ S_{n,\mathcal{X}^A,\mathcal{X}_3} \sqrt{m} \left(\tilde{A} - A \right)' \beta_2 + O_p\left(m^{-1/2}\right) \right\}, \qquad (6)
\end{aligned}$$

where, by (2),

$$\begin{aligned}
\sqrt{m} \left(\tilde{A} - A \right)' \beta_2 &= \left(\frac{1}{m} \sum_{j=1}^{m} \mathcal{X}_{3j} \mathcal{X}_{3j}' \right)^{-1} \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \mathcal{X}_{3j} e_j' \beta_2 \\
&\xrightarrow{d} N\left(0_{(d_3+1)\times 1}, \Phi^{-1}_{\mathcal{X}_3,\mathcal{X}_3} E\left(\mathcal{X}_3 \mathcal{X}_3' \beta_2' e e' \beta_2\right) \Phi^{-1}_{\mathcal{X}_3,\mathcal{X}_3}\right).
\end{aligned}$$

Therefore,

$$S_{n,\mathcal{X}^A,\mathcal{X}_3}\sqrt{m}\left(\tilde{A}-A\right)'\beta_2 \xrightarrow{d} N\left(0_{(d+1)\times 1},\Psi_2\right), \tag{7}$$

where

$$\Psi_2 = \Phi_{\mathcal{X}^A,\mathcal{X}_3}\Phi^{-1}_{\mathcal{X}_3,\mathcal{X}_3}E\left(\mathcal{X}_3\mathcal{X}_3'\beta_2'ee'\beta_2\right)\Phi^{-1}_{\mathcal{X}_3,\mathcal{X}_3}\Phi'_{\mathcal{X}^A,\mathcal{X}_3}.$$

Finally, by $\tilde{A} \perp\!\!\!\perp \mathcal{S}_1$ and $E\left(\tilde{A}\right)=A$,

$$E\left[\left(\mathcal{X}_i^A v_i\right)\left\{\mathcal{X}_i^A\mathcal{X}_{3i}'\left(\tilde{A}-A\right)'\beta_2\right\}'\right] = E\left(\mathcal{X}^A\mathcal{X}^{A\prime}v\mathcal{X}_3'\right)E\left(\tilde{A}-A\right)'\beta_2 = 0_{(d+1)\times(d+1)} \tag{8}$$

so that two CLT terms (5) and (6) are asymptotically independent.

In what follows, we consider three different divergence scenarios of (n,m). The asymptotic analysis in each scenario is based on substituting (4), (5), (7), and (8) into (3). First, if $n/m \to \kappa \in (0,\infty)$, then

$$\begin{aligned}
&\sqrt{n}\left(\hat{\beta}_{PARA}-\beta\right)\\
&= S^{-1}_{n,\tilde{X},\tilde{X}}\left(\sqrt{n}S_{n,\tilde{X},v}-\sqrt{n}S_{n,\tilde{X},\mathcal{X}_3'\left(\tilde{A}-A\right)'\beta_2}\right)\\
&= \left\{S_{n,\mathcal{X}^A,\mathcal{X}^A}+o_p(1)\right\}^{-1}\left\{\sqrt{n}S_{n,\mathcal{X}^A,v}-\sqrt{\frac{n}{m}}S_{n,\mathcal{X}^A,\mathcal{X}_3}\sqrt{m}\left(\tilde{A}-A\right)'\beta_2+o_p(1)\right\}\\
&\xrightarrow{d} N\left(0_{(d+1)\times 1},\Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\Psi\Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\right),
\end{aligned}$$

where $\Psi=\Psi_1+\kappa\Psi_2$. Second, if $n/m\to 0$, then

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}_{PARA}-\beta\right) &= S^{-1}_{n,\tilde{X},\tilde{X}}\left(\sqrt{n}S_{n,\tilde{X},v}-\sqrt{n}S_{n,\tilde{X},\mathcal{X}_3'\left(\tilde{A}-A\right)'\beta_2}\right)\\
&= \left\{S_{n,\mathcal{X}^A,\mathcal{X}^A}+o_p(1)\right\}^{-1}\left\{\sqrt{n}S_{n,\mathcal{X}^A,v}+o_p(1)\right\}\\
&\xrightarrow{d} N\left(0_{(d+1)\times 1},\Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\Psi_1\Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\right).
\end{aligned}$$

Third, if $n/m \to \infty$, then

$$\begin{aligned}
&\sqrt{m}\left(\hat{\beta}_{PARA} - \beta\right) \\
&= S^{-1}_{n,\tilde{X},\tilde{X}}\left(\sqrt{m}S_{n,\tilde{X},v} - \sqrt{m}S_{n,\tilde{X},\mathcal{X}_3'(\tilde{A}-A)'\beta_2}\right) \\
&= \left\{S_{n,\mathcal{X}^A,\mathcal{X}^A} + o_p(1)\right\}^{-1}\left[\sqrt{\frac{m}{n}}\left\{\sqrt{n}S_{n,\mathcal{X}^A,v} + o_p(1)\right\} + S_{n,\mathcal{X}^A,\mathcal{X}_3}\sqrt{m}\left(\tilde{A} - A\right)'\beta_2 + O_p\left(m^{-1/2}\right)\right] \\
&= \left\{S_{n,\mathcal{X}^A,\mathcal{X}^A} + o_p(1)\right\}^{-1}\left\{S_{n,\mathcal{X}^A,\mathcal{X}_3}\sqrt{m}\left(\tilde{A} - A\right)'\beta_2 + o_p(1)\right\} \\
&\xrightarrow{d} N\left(0_{(d+1)\times 1}, \Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\Psi_1\Phi^{-1}_{\mathcal{X}^A,\mathcal{X}^A}\right).
\end{aligned}$$

Lastly, it is straightforward to demonstrate consistency of $\hat{\Psi}_1$, $\hat{\Psi}_2$ and $\hat{\Psi} = \hat{\Psi}_1 + (n/m)\,\hat{\Psi}_2$ for Ψ_1, Ψ_2 and Ψ, respectively. Therefore, the proof for this part is omitted.

■

1.3 MSOLS, MSII and MSII-FM Estimators

Hirukawa and Prokhorov (2018) propose a set of two-sample estimators to address imputation biases caused by a practice known as "hot-deck imputation" in Census. Missing values for non-respondents are imputed using the values from respondents with similar characteristics without accounting for the imputation bias. Effectively, the MSII estimator exploits the structure of a surrogate sample obtained by matching on variables that are common to both respondents and non-respondents.

The MSII estimation starts from obtaining a combined sample by means of NNM on the observed variables. As before, let $\mathcal{S}_1 = (Y, X_1, X_3)$ denote a sample for which we have all responses and let $\mathcal{S}_2 = (X_2, X_3)$ denote the sample which contains the variable with non-response as well as the observed variables to match on. Hot-deck imputation uses a distance measure between X_3 in $\mathcal{S}_1$ and X_3 in $\mathcal{S}_2$ to obtain a matched value of X_2 for $\mathcal{S}_1$. Specifically, for each observation of X_3 in $\mathcal{S}_1$, K-NNM picks out K closest matches of X_2 from $\mathcal{S}_2$ through finding K closest matches of X_3

in S_2 with respect to the Mahalanobis distance. The resulting combined sample can be written as

$$\mathcal{S} = \left\{ \left(Y_i, X_{1i}, X_{2j_1(i)}, \ldots, X_{2j_K(i)}, X_{3Ii}, X_{3Ei} \right) \right\}_{i=1}^{n}.$$

The MSOLS estimator for the regression of Y_i on $X_{i,j(i)} := \left(1, X'_{1i}, X'_{2j(i)}, X'_{3Ii}\right)'$, where $X_{2j(i)} := (1/K) \sum_{k=1}^{K} X_{2j_k(i)}$, generates a non-vanishing, classical measurement error (or attenuation) bias. The bias is attributed to using $X_{2j(i)}$ as a proxy for X_{2i}, and the source of attenuation is $\Sigma_2 = E\left(\eta_2 \eta_2'\right)$.

As is well known in the literature on errors-in-variables models, the bias cannot be corrected in general without imposing additional identification conditions. However, in the above two-sample setup it can be corrected analytically with no such extra conditions because $\mathcal{S}_2$ serves as repeated measurements. MSII is a bias-corrected estimator defined as

$$\hat{\beta}_{II} := \left(\frac{1}{n} \sum_{i=1}^{n} X_{i,j(i)} X'_{i,j(i)} - \frac{1}{K} \hat{\Sigma} \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} X_{i,j(i)} Y_i,$$

where $\hat{\Sigma} := \operatorname{diag}\left\{ 0_{(d_1+1)\times(d_1+1)}, \hat{\Sigma}_2, 0_{d_{3I}\times d_{3I}} \right\}$, and $\hat{\Sigma}_2$ is the difference-based variance estimator that can be obtained by reordering $\mathcal{S}_2$ with respect to X_3.

Notice that MSII can attain $\sqrt{n}$-asymptotic normality when the number of matching variables is one. To overcome the curse of dimensionality in continuous matching variables, Hirukawa and Prokhorov (2018) propose a two-step estimator called the fully-modified MSII (MSII-FM) estimator. In its second step, MSII-FM eliminates the second-order bias due to the so-called *matching discrepancy* asymptotically by means of a polynomial approximation similar to the one studied by Abadie and Imbens (2011) in the context of average treatment effect estimation. It is demonstrated that MSII-FM can achieve parametric convergence when the number of matching variables is no greater than three.

2 Additional Monte Carlo Results

This section presents the results from four different simulation designs. First two studies deal with the cases of both empty and non-empty X_1. The difference is that while the first study considers only linear instruments, the second one covers both linear and nonlinear instruments. The third study is based on a somewhat different design with a wider range of functions $h(x)$ used in the reduced form, and focuses exclusively on the cases of non-empty X_1. The fourth study is similar to the one included in the paper, with the difference that X_1 is an independent regressor. The results below indicate superior performances of PILS.

2.1 Simulation Study I

The first simulation study covers the cases of both empty and non-empty X_1. We begin with the case of non-empty X_1 and explain how the data are generated. The long regression is in the form of

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_{3I_C} + u, \tag{9}$$

where true parameter values are $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$, and β_3 is the parameter of interest. The short regression after omitting X_2 from (9) is

$$Y = \beta_0 + \beta_1 X_1 + \beta_3 X_{3I_C} + (u + \beta_2 X_2) =: \beta_0 + \beta_1 X_1 + \beta_3 X_{3I_C} + v. \tag{10}$$

There are three common variables, namely, an included continuous one X_{3I_C}, an excluded continuous one X_{3E_C} and an excluded discrete one X_{3E_D}. Also for $r \geq 0$ and some function $h(\cdot)$ specified shortly, let $H_r := E\left\{X_{3I_C}^r h\left(X_{3I_C}\right)\right\}$. Then, X_1, X_2

and Z, the instrument for X_{3I_C}, are generated, respectively, as

$$\begin{aligned} X_1 &= 1 + X_{3I_C}^2 + X_{3E_C} + 2X_{3E_D} + \eta_1, \\ X_2 &= h\left(X_{3I_C}\right) - \frac{3}{4}\left(H_2 + 1\right) X_{3E_C} + 2X_{3E_D} + \eta_2, \text{ and} \quad (11) \\ Z &= \pi\left(X_{3I_C} - 2H_1 X_{3E_D}\right) + \xi, \quad (12) \end{aligned}$$

where $\pi = \sqrt{3\rho^2 / \left\{4 - \rho^2\left(4 + 3H_1^2\right)\right\}}$ and $\rho = corr\left(X_{3I_C}, Z\right) \in \{0.1, 0.4\}$. Notice that $\rho = 0.4$ roughly mimics the sample correlation coefficient between the endogenous regressor (*educ*) and its IV (*fatheduc*) in Section 4 of the paper.

When X_1 is empty in (9), we consider two alternative cases for the pair of the missing regressor X_2 and the instrument Z. The first case deals with an *additively separable* reduced form of X_2. In this case, the pair of X_2 and Z are still (11) and (12), respectively. The second case is based on an *additively non-separable* reduced form of X_2. In this case, X_2 and Z are specified as

$$\begin{aligned} X_2 &= h\left(X_{3I_C}\right) X_{3E_C}^2 + 2X_{3E_D} + \eta_2, \text{ and} \\ Z &= \zeta\left(X_{3I_C} - \frac{8}{3} H_1 X_{3E_D}\right) + \xi, \end{aligned}$$

where $\zeta = \sqrt{9\rho^2 / \left\{12 - \rho^2\left(12 + 16H_1^2\right)\right\}}$.

For each case, the function $h(\cdot)$ is specified as one of the following:

$$h(x) = \begin{cases} x & \text{[Model A]} \\ \sin(\pi x/4) & \text{[Model B]} \\ \{(5/3)x + 3/2\}\mathbf{1}\{x < 0\} + \{-(4/3)x + 3/2\}\mathbf{1}\{x \geq 0\} & \text{[Model C]} \\ x + (5/\tau)\phi(x/\tau) - (5/2)\{\Phi(2/\tau) - 1/2\},\ \tau = 3/4 & \text{[Model D]} \end{cases},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the pdf and cdf of $N(0,1)$, respectively. Models A and B are monotone. Linearity of $h(\cdot)$ in Model A does not cause identification failure, because two common variables $\left(X_{3E_C}, X_{3E_D}\right)$ are excluded from (9). Model B is nonlinear and odd. Models C and D are non-monotone, and the former is even non-smooth when $x = 0$. Inspired by the Monte Carlo design of Horowitz and Spokoiny (2001), Model D can be viewed as a linear function with a bump.

The following distributional assumptions also apply.

$$\begin{aligned}
&(u, \eta_1, \eta_2, \xi)' \sim N(0_{4\times 1}, I_4).\\
&X_{3I_C}, X_{3E_C} \overset{iid}{\sim} U[-2, 2].\\
&X_{3E_D} \sim Bernoulli\,(1/2) - 1/2 = \begin{cases} -1/2 & \text{w.p. } 1/2 \\ \;\;1/2 & \text{w.p. } 1/2 \end{cases}.\\
&u \perp\!\!\!\perp X_3 = (X_{3I_C}, X_{3E_C}, X_{3E_D}).
\end{aligned}$$

The distributional assumption also implies that values of (H_1, H_2, H_3) are:

$$(H_1, H_2, H_3) = \begin{cases} (0, 4/3, 0) & \text{[Model A]} \\ (0, 8/\pi^2, 0) & \text{[Model B]} \\ (0, 2/9, -1) & \text{[Model C]} \\ (0, 4/3, -0.9989) & \text{[Model D]} \end{cases}.$$

The above procedure provides us with two observable samples

$$\mathcal{S}_1 = \{(Y_i, X_{1i}, X_{3I_C i}, X_{3E_C i}, X_{3E_D i}, Z_i)\}_{i=1}^n \text{ and } \mathcal{S}_2 = \{(X_{2j}, X_{3I_C j}, X_{3E_C j}, X_{3E_D j})\}_{j=1}^m.$$

The complete sample

$$\mathcal{S}^* = \{(Y_i, X_{1i}, X_{2i}, X_{3I_C i}, X_{3E_C i}, X_{3E_D i}, Z_i)\}_{i=1}^n$$

is the sample that would not be observed in practice. Finally, the matched sample

$$\mathcal{S} = \left\{\left(Y_i, X_{1i}, X_{2j_1(i)}, \ldots, X_{2j_K(i)}, X_{3I_C i}, X_{3E_C i}, X_{3E_D i}, Z_i\right)\right\}_{i=1}^n$$

is constructed via the NNM with respect to $X_3 = (X_{3I_C}, X_{3E_C}, X_{3E_D})$ as in Hirukawa and Prokhorov (2018). The NNM is based on the Mahalanobis distance, and the number of matches is set equal to $K = 1$ (single match) that is most commonly chosen. Sample sizes of $\mathcal{S}_1$ and $\mathcal{S}_2$ are $n \in \{500, 1000, 2000\}$ and $m = n/2$. For each combination of the functional form of $h(\cdot)$ and value of π, we draw 1000 Monte Carlo samples.

We compare the following estimators of β_3: (i) the infeasible OLS estimator for (9) using $\mathcal{S}^*$ [OLS*]; (ii) the inconsistent OLS estimator for (10) using $\mathcal{S}_1$ only [OLS-S]; (iii) the IV estimator for (10) using $\mathcal{S}_1$ only [IV-S]; (iv) the inconsistent MSOLS estimator for (9) using $\mathcal{S}$ [MSOLS]; (v) the MSII-FM estimator for (9) using $\mathcal{S}$ and a second-order polynomial [MSII-FM], along with the first-step MSII estimator [MSII]; (vi) the PARA estimator for (9) using $\mathcal{S}_1$ and $\mathcal{S}_2$ [PARA]; (vii) the PILS estimator for (9) using $\mathcal{S}_1$, $\mathcal{S}_2$ and the Epanechnikov kernel [PILS-E]; and (viii) the PILS estimator for (9) using $\mathcal{S}_1$, $\mathcal{S}_2$ and the beta kernel [PILS-B].

Notice that OLS* is consistent, because $E(u) = E(X_2 u) = E(X_{3I_C} u) = 0$. Moreover, $E(v) = 0$, $E(X_{3I_C} v) \neq 0$, $\rho = corr(X_{3I_C}, Z) \neq 0$ and $E(Zv) = 0$, which implies inconsistency of OLS-S and consistency of IV-S.

When X_1 is empty in (9), PARA is consistent by Proposition S1. On the other hand, when X_1 is non-empty, the functional form of the imputed regressor in PARA matches the reduced form of X_2 for Model A only. In reality, however, even if the reduced form of X_2 is misspecified, it is still possible to estimate the regression (9) consistently by PARA. It turns out that under this Monte Carlo design, PARA is consistent if and only if $h(\cdot)$ is odd, i.e., for both Models A and B, as indicated in Proposition S2.

Implementing the two PILS estimators requires a choice of the smoothing parameters. We put $\hat{h} = \hat{\sigma}_X (\log m/m)^{0.3}$ for the Epanechnikov kernel, $\hat{b} = \hat{\sigma}_U (\log m/m)^{0.6}$ for the beta kernel, and $\hat{\lambda} = (\log m/m)^{0.6}$ for the discrete kernel, where $\hat{\sigma}_X$ and $\hat{\sigma}_U$ are sample standard deviations of the continuous common variable in the original scale $X (= X_{3I_C}, X_{3E_C})$ and in the transformed scale $U := (X + 2)/4 \in [0, 1]$, respectively. These shrinkage rates indeed fulfill the requirements in Section 2.4.4 of the paper.

For each estimator of β_3, the following performance measures are computed: (i) $Mean$ (simulation average of the parameter estimate); (ii) SD (simulation standard deviation of the parameter estimate); (iii) $RMSE$ (root mean-squared error of the parameter estimate); (iv) $\overline{SE}$ (simulation average of the standard error); and (v)

CR (coverage rate for the nominal 95% confidence interval). Formulae for standard errors are as follows: (a) heteroskedasticity-robust standard errors by Eicker (1963) and White (1980) are calculated for OLS* and IV-S; (b) the one for MSII-FM follows Theorem 4 of Hirukawa and Prokhorov (2018); (c) the one for PILS appears in Theorem 2; and (d) the one for PARA is given in Proposition S2 above. Because of inconsistency we do not compute $\overline{SE}$ or CR of OLS-S and MSOLS. Those of the first-step MSII are also omitted due to its nonparametric convergence rate.

TABLES 1-3 HERE

Table 1 presents comprehensive results when X_1 is non-empty. The results when X_1 is empty are available in Tables 2 (when X_2 has an additively separable reduced form) and 3 (when X_2 has a non-separable reduced form). It can be found that PILS-B performs well regardless of the models.

2.2 Simulation Study II

In this simulation study, the long and short regression models are still (9) and (10), respectively. As before, the true parameter values are $\beta_0 = \beta_1 = \beta_2 = \beta_3 = 1$, and β_3 is assumed to be the parameter of interest. The regressor X_1 either enters or is excluded from the regression.

Table 1: Simulation Study I: Estimates of β_3 in Presence of X_1

	$(n,m)=(500,250)$					$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
Estimator	$Mean$	SD	$RMSE$	$\overline{SE}$	CR	$Mean$	SD	$RMSE$	$\overline{SE}$	CR	$Mean$	SD	$RMSE$	$\overline{SE}$	CR
Model A: $h(x)=x$															
OLS*	0.9969	0.0483	0.0484	0.0470	94%	0.9998	0.0336	0.0336	0.0334	95%	0.9998	0.0236	0.0236	0.0235	95%
OLS-S	1.9963	0.0746	0.9991	–	–	2.0004	0.0516	1.0017	–	–	1.9997	0.0383	1.0004	–	–
IV-S: $\rho=0.4$	0.9968	0.2050	0.2051	0.2031	94%	1.0004	0.1460	0.1460	0.1431	95%	1.0042	0.1019	0.1020	0.1008	94%
$\rho=0.1$	0.5981	5.0285	5.0445	14.6315	96%	0.9028	0.8699	0.8754	0.8039	95%	0.9376	0.5032	0.5070	0.4897	96%
MSOLS	1.3733	0.0910	0.3843	–	–	1.3686	0.0639	0.3741	–	–	1.3684	0.0441	0.3711	–	–
MSII	0.9260	0.1669	0.1825	–	–	0.9671	0.1105	0.1153	–	–	0.9868	0.0714	0.0726	–	–
MSII-FM	0.8915	0.1733	0.2045	0.1528	91%	0.9518	0.1125	0.1224	0.0972	91%	0.9795	0.0719	0.0748	0.0656	93%
PARA	0.9959	0.1055	0.1056	0.1029	95%	1.0004	0.0727	0.0727	0.0725	95%	1.0035	0.0533	0.0534	0.0514	94%
PILS-E	1.1494	0.1177	0.1902	0.1186	73%	1.0847	0.0773	0.1147	0.0881	85%	1.0569	0.0541	0.0785	0.0650	89%
PILS-B	1.0092	0.1171	0.1175	0.1188	95%	1.0080	0.0769	0.0773	0.0815	96%	1.0108	0.0549	0.0559	0.0602	96%
Model B: $h(x)=\sin(\pi x/4)$															
OLS*	0.9968	0.0428	0.0429	0.0419	94%	0.9996	0.0295	0.0295	0.0297	95%	1.0000	0.0212	0.0212	0.0210	95%
OLS-S	1.6045	0.0746	0.6090	–	–	1.6084	0.0517	0.6106	–	–	1.6075	0.0383	0.6087	–	–
IV-S: $\rho=0.4$	0.9977	0.1950	0.1950	0.1939	95%	1.0023	0.1391	0.1391	0.1367	96%	1.0045	0.0973	0.0974	0.0965	95%
$\rho=0.1$	0.8366	2.7156	2.7205	5.9106	98%	0.9459	0.7678	0.7697	0.7171	96%	0.9591	0.4442	0.4461	0.4417	97%
MSOLS	1.2200	0.0816	0.2346	–	–	1.2205	0.0573	0.2278	–	–	1.2227	0.0399	0.2263	–	–
MSII	0.9553	0.1252	0.1330	–	–	0.9800	0.0860	0.0883	–	–	0.9937	0.0565	0.0568	–	–
MSII-FM	0.9404	0.1278	0.1410	0.1024	88%	0.9733	0.0869	0.0909	0.0678	86%	0.9906	0.0567	0.0575	0.0465	89%
PARA	0.9957	0.0898	0.0899	0.0880	95%	0.9994	0.0620	0.0620	0.0621	95%	1.0029	0.0455	0.0456	0.0440	93%
PILS-E	1.0832	0.1022	0.1318	0.0969	84%	1.0468	0.0674	0.0820	0.0717	91%	1.0330	0.0465	0.0571	0.0529	93%
PILS-B	0.9930	0.1012	0.1014	0.0962	93%	0.9961	0.0668	0.0669	0.0664	96%	1.0019	0.0470	0.0471	0.0490	96%
Model C: $h(x)=\{(5/3)x+3/2\}\mathbf{1}\{x<0\}+\{-(4/3)x+3/2\}\mathbf{1}\{x\geq 0\}$															
OLS*	0.9967	0.0393	0.0394	0.0389	96%	0.9996	0.0266	0.0266	0.0276	96%	1.0001	0.0199	0.0199	0.0195	94%
OLS-S	1.1631	0.0783	0.1809	–	–	1.1658	0.0520	0.1738	–	–	1.1667	0.0384	0.1711	–	–
IV-S: $\rho=0.4$	0.9991	0.2003	0.2003	0.1893	94%	1.0002	0.1314	0.1314	0.1334	95%	1.0041	0.0912	0.0913	0.0943	96%
$\rho=0.1$	1.0255	3.4778	3.4779	5.6981	99%	0.9855	0.6594	0.6595	0.6506	98%	1.0076	0.3991	0.3992	0.4120	97%
MSOLS	1.0591	0.0766	0.0968	–	–	1.0601	0.0527	0.0799	–	–	1.0630	0.0381	0.0736	–	–
MSII	0.9828	0.1024	0.1038	–	–	0.9913	0.0695	0.0700	–	–	0.9996	0.0472	0.0472	–	–
MSII-FM	0.9799	0.1022	0.1042	0.0737	85%	0.9896	0.0699	0.0706	0.0502	84%	0.9988	0.0473	0.0473	0.0350	85%
PARA	0.9508	0.1186	0.1284	–	–	0.9561	0.0825	0.0935	–	–	0.9586	0.0594	0.0724	–	–
PILS-E	1.0215	0.0926	0.0951	0.0944	95%	1.0111	0.0607	0.0617	0.0710	97%	1.0104	0.0419	0.0432	0.0532	98%
PILS-B	0.9946	0.0920	0.0922	0.0972	96%	0.9959	0.0604	0.0605	0.0664	97%	1.0008	0.0420	0.0420	0.0495	98%
Model D: $h(x)=x+(5/\tau)\phi(x/\tau)-(5/2)\{\Phi(2/\tau)-1/2\},\ \tau=3/4$															
OLS*	0.9974	0.0481	0.0481	0.0468	94%	1.0004	0.0325	0.0325	0.0332	95%	0.9997	0.0233	0.0233	0.0235	95%
OLS-S	1.9967	0.0773	0.9997	–	–	1.9992	0.0513	1.0005	–	–	2.0000	0.0379	1.0008	–	–
IV-S: $\rho=0.4$	0.9970	0.2142	0.2142	0.2034	94%	0.9966	0.1419	0.1419	0.1432	96%	1.0038	0.1013	0.1013	0.1008	94%
$\rho=0.1$	0.4076	12.3720	12.3861	36.0469	97%	0.8913	0.8477	0.8547	0.8034	95%	0.9587	0.4982	0.4999	0.4896	96%
MSOLS	1.3717	0.0913	0.3828	–	–	1.3634	0.0624	0.3687	–	–	1.3616	0.0445	0.3644	–	–
MSII	0.9104	0.1673	0.1898	–	–	0.9556	0.1089	0.1176	–	–	0.9803	0.0682	0.0710	–	–
MSII-FM	0.8758	0.1740	0.2138	0.1573	91%	0.9397	0.1113	0.1266	0.0985	90%	0.9730	0.0689	0.0740	0.0655	94%
PARA	0.7282	0.1876	0.3303	–	–	0.7369	0.1320	0.2944	–	–	0.7371	0.0948	0.2795	–	–
PILS-E	1.1442	0.1166	0.1854	0.1189	75%	1.0774	0.0752	0.1079	0.0884	89%	1.0510	0.0519	0.0728	0.0652	91%
PILS-B	0.9873	0.1184	0.1190	0.1223	95%	0.9884	0.0761	0.0770	0.0830	96%	0.9959	0.0529	0.0531	0.0609	98%

Notes: $Mean$ = simulation average of the parameter estimate; SD = simulation standard deviation of the parameter estimate; $RMSE$ = root mean-squared error of the parameter estimate; $RMSE$ = root mean-squared error of the parameter estimate; $\overline{SE}$ = simulation average of the standard error; and CR = coverage rate for the nominal 95% confidence interval.

Table 2: Simulation Study I: Estimates of β_3 in Absence of X_1 (Separable Case)

Estimator	$(n,m)=(500,250)$					$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
	$Mean$	SD	$RMSE$	$\overline{SE}$	CR	$Mean$	SD	$RMSE$	$\overline{SE}$	CR	$Mean$	SD	$RMSE$	$\overline{SE}$	CR
Model A: $h(x)=x$															
OLS*	1.0022	0.0459	0.0460	0.0472	96%	1.0011	0.0334	0.0335	0.0334	95%	1.0013	0.0241	0.0242	0.0236	95%
OLS-S	2.0011	0.0735	1.0038	–	–	2.0019	0.0522	1.0033	–	–	2.0022	0.0369	1.0029	–	–
IV-S: $\rho=0.4$	0.9980	0.2016	0.2016	0.2046	95%	0.9964	0.1463	0.1463	0.1450	95%	1.0018	0.1038	0.1039	0.1019	94%
$\rho=0.1$	0.5895	4.9621	4.9791	9.1768	95%	0.8575	0.9308	0.9416	0.8960	95%	0.9257	0.5001	0.5056	0.4920	96%
MSOLS	1.3833	0.0897	0.3936	–	–	1.3711	0.0637	0.3765	–	–	1.3652	0.0456	0.3680	–	–
MSII	0.9474	0.1638	0.1720	–	–	0.9644	0.1093	0.1150	–	–	0.9775	0.0732	0.0766	–	–
MSII-FM	0.9134	0.1709	0.1916	0.1488	92%	0.9488	0.1107	0.1220	0.0979	92%	0.9700	0.0737	0.0795	0.0663	92%
PARA	1.0111	0.1069	0.1075	0.1023	94%	1.0008	0.0749	0.0749	0.0727	94%	0.9988	0.0517	0.0517	0.0515	95%
PILS-E	1.1722	0.1143	0.2067	0.0947	53%	1.0893	0.0786	0.1190	0.0680	70%	1.0538	0.0531	0.0756	0.0490	77%
PILS-B	1.0276	0.1144	0.1177	0.0988	90%	1.0112	0.0789	0.0797	0.0706	91%	1.0068	0.0531	0.0535	0.0505	94%
Model B: $h(x)=\sin(\pi x/4)$															
OLS*	1.0019	0.0409	0.0410	0.0420	96%	1.0014	0.0297	0.0297	0.0297	96%	1.0010	0.0215	0.0215	0.0210	95%
OLS-S	1.6091	0.0734	0.6135	–	–	1.6099	0.0522	0.6121	–	–	1.6100	0.0370	0.6112	–	–
IV-S: $\rho=0.4$	0.9983	0.1894	0.1894	0.1942	96%	0.9996	0.1378	0.1378	0.1377	95%	1.0044	0.0991	0.0992	0.0969	94%
$\rho=0.1$	0.9333	4.3619	4.3624	8.1980	97%	0.9261	0.7885	0.7920	0.7627	97%	0.9512	0.4429	0.4456	0.4428	97%
MSOLS	1.2295	0.0807	0.2433	–	–	1.2235	0.0570	0.2307	–	–	1.2198	0.0409	0.2236	–	–
MSII	0.9727	0.1243	0.1272	–	–	0.9808	0.0850	0.0871	–	–	0.9862	0.0577	0.0593	–	–
MSII-FM	0.9577	0.1273	0.1341	0.1001	88%	0.9738	0.0856	0.0895	0.0680	88%	0.9830	0.0579	0.0603	0.0469	88%
PARA	1.0087	0.0906	0.0911	0.0876	94%	1.0014	0.0635	0.0635	0.0622	94%	0.9987	0.0443	0.0444	0.0441	94%
PILS-E	1.1003	0.0990	0.1409	0.0786	69%	1.0511	0.0680	0.0850	0.0559	79%	1.0297	0.0459	0.0547	0.0399	84%
PILS-B	1.0074	0.0988	0.0991	0.0805	89%	0.9998	0.0679	0.0679	0.0574	91%	0.9982	0.0457	0.0457	0.0409	92%
Model C: $h(x)=\{(5/3)x+3/2\}\mathbf{1}\{x<0\}+\{-(4/3)x+3/2\}\mathbf{1}\{x\geq 0\}$															
OLS*	1.0014	0.0383	0.0383	0.0389	96%	1.0017	0.0274	0.0274	0.0276	95%	1.0006	0.0200	0.0200	0.0195	95%
OLS-S	1.1659	0.0748	0.1820	–	–	1.1681	0.0535	0.1764	–	–	1.1684	0.0376	0.1726	–	–
IV-S: $\rho=0.4$	0.9953	0.1858	0.1859	0.1889	97%	1.0001	0.1308	0.1308	0.1341	95%	1.0037	0.0964	0.0965	0.0945	94%
$\rho=0.1$	0.0295	26.8821	26.8996	148.4350	99%	1.0158	0.7846	0.7847	0.7344	98%	0.9839	0.4051	0.4054	0.4121	98%
MSOLS	1.0677	0.0759	0.1018	–	–	1.0637	0.0534	0.0832	–	–	1.0601	0.0380	0.0711	–	–
MSII	0.9954	0.1019	0.1020	–	–	0.9953	0.0695	0.0697	–	–	0.9935	0.0483	0.0487	–	–
MSII-FM	0.9894	0.1018	0.1024	0.0732	85%	0.9939	0.0693	0.0696	0.0503	86%	0.9928	0.0483	0.0488	0.0351	85%
PARA	1.0041	0.1027	0.1028	0.1021	94%	1.0005	0.0722	0.0722	0.0723	96%	0.9976	0.0513	0.0513	0.0511	94%
PILS-E	1.0305	0.0896	0.0947	0.0732	87%	1.0148	0.0604	0.0622	0.0524	90%	1.0069	0.0415	0.0421	0.0377	93%
PILS-B	1.0025	0.0892	0.0893	0.0779	91%	1.0002	0.0602	0.0602	0.0552	93%	0.9972	0.0414	0.0415	0.0392	93%
Model D: $h(x)=x+(5/\tau)\phi(x/\tau)-(5/2)\{\Phi(2/\tau)-1/2\},\ \tau=3/4$															
OLS*	1.0011	0.0466	0.0466	0.0469	96%	1.0017	0.0332	0.0333	0.0332	95%	1.0010	0.0243	0.0243	0.0235	95%
OLS-S	1.9991	0.0739	1.0018	–	–	2.0013	0.0529	1.0027	–	–	2.0017	0.0372	1.0024	–	–
IV-S: $\rho=0.4$	0.9951	0.2060	0.2061	0.2062	95%	0.9938	0.1404	0.1405	0.1460	96%	1.0003	0.1029	0.1029	0.1026	95%
$\rho=0.1$	0.5677	5.3440	5.3615	9.8155	96%	0.8730	0.9114	0.9202	0.8826	95%	0.9335	0.4957	0.5001	0.4930	96%
MSOLS	1.3786	0.0906	0.3893	–	–	1.3661	0.0633	0.3715	–	–	1.3577	0.0454	0.3606	–	–
MSII	0.9218	0.1688	0.1860	–	–	0.9554	0.1086	0.1175	–	–	0.9701	0.0725	0.0784	–	–
MSII-FM	0.8837	0.1752	0.2102	0.1560	92%	0.9399	0.1105	0.1258	0.0987	91%	0.9628	0.0730	0.0819	0.0662	90%
PARA	1.0097	0.1562	0.1565	0.1505	93%	1.0003	0.1085	0.1085	0.1066	95%	0.9965	0.0769	0.0770	0.0754	95%
PILS-E	1.1596	0.1136	0.1959	0.0952	59%	1.0840	0.0764	0.1135	0.0681	73%	1.0477	0.0520	0.0705	0.0491	81%
PILS-B	0.9974	0.1157	0.1157	0.1025	91%	0.9942	0.0781	0.0784	0.0722	93%	0.9917	0.0527	0.0533	0.0513	95%

Notes: $Mean$ = simulation average of the parameter estimate; SD = simulation standard deviation of the parameter estimate; $RMSE$ = root mean-squared error of the parameter estimate; $\overline{SE}$ = simulation average of the standard error; and CR = coverage rate for the nominal 95% confidence interval.

Table 3: Simulation Study I: Estimates of β_3 in Absence of X_1 (Non-Separable Case)

	$(n,m)=(500,250)$					$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
Estimator	$Mean$	SD	$RMSE$	$\overline{SE}$	CR	$Mean$	SD	$RMSE$	$\overline{SE}$	CR	$Mean$	SD	$RMSE$	$\overline{SE}$	CR
Model A: $h(x)=x$															
OLS*	1.0014	0.0488	0.0488	0.0490	95%	1.0019	0.0341	0.0342	0.0347	95%	1.0007	0.0247	0.0247	0.0246	95%
OLS-S	2.3290	0.0966	1.3325	–	–	2.3363	0.0676	1.3380	–	–	2.3356	0.0474	1.3364	–	–
IV-S: $\rho=0.4$	0.9860	0.2389	0.2393	0.2452	95%	0.9932	0.1753	0.1754	0.1733	95%	0.9980	0.1214	0.1214	0.1221	95%
$\rho=0.1$	−0.8925	46.7061	46.7444	742.6379	95%	0.7964	1.1863	1.2036	1.1256	95%	0.9083	0.5985	0.6054	0.5924	96%
MSOLS	1.4108	0.1038	0.4237	–	–	1.3748	0.0710	0.3815	–	–	1.3542	0.0512	0.3579	–	–
MSII	0.7994	0.2095	0.2901	–	–	0.8819	0.1226	0.1702	–	–	0.9334	0.0769	0.1018	–	–
MSII-FM	0.7450	0.2215	0.3378	0.2107	86%	0.8592	0.1251	0.1884	0.1175	82%	0.9228	0.0776	0.1094	0.0721	82%
PARA	1.0079	0.2428	0.2429	0.2385	95%	1.0064	0.1733	0.1734	0.1674	94%	1.0003	0.1163	0.1163	0.1189	95%
PILS-E	1.1453	0.1278	0.1935	0.1140	71%	1.0630	0.0814	0.1030	0.0827	88%	1.0303	0.0565	0.0641	0.0602	94%
PILS-B	1.0038	0.1251	0.1252	0.1225	94%	0.9897	0.0824	0.0831	0.0873	96%	0.9871	0.0569	0.0583	0.0621	96%
Model B: $h(x)=\sin(\pi x/4)$															
OLS*	1.0016	0.0442	0.0443	0.0445	95%	1.0018	0.0309	0.0310	0.0315	95%	1.0007	0.0226	0.0226	0.0223	95%
OLS-S	1.8088	0.0769	0.8124	–	–	1.8132	0.0540	0.8150	–	–	1.8125	0.0382	0.8134	–	–
IV-S: $\rho=0.4$	0.9899	0.1937	0.1940	0.1996	95%	0.9980	0.1395	0.1395	0.1413	95%	1.0005	0.1000	0.1000	0.0996	95%
$\rho=0.1$	0.7602	3.5808	3.5888	5.9520	97%	0.8916	0.8264	0.8335	0.8225	96%	0.9394	0.4610	0.4650	0.4646	96%
MSOLS	1.3230	0.0870	0.3345	–	–	1.3099	0.0602	0.3157	–	–	1.3006	0.0434	0.3037	–	–
MSII	0.9168	0.1602	0.1806	–	–	0.9486	0.1036	0.1157	–	–	0.9687	0.0673	0.0742	–	–
MSII-FM	0.8922	0.1662	0.1981	0.1428	91%	0.9379	0.1051	0.1220	0.0893	89%	0.9636	0.0677	0.0769	0.0587	88%
PARA	1.0051	0.1435	0.1436	0.1394	94%	1.0037	0.1018	0.1019	0.0983	94%	0.9987	0.0693	0.0693	0.0699	95%
PILS-E	1.1378	0.1086	0.1754	0.0863	61%	1.0689	0.0720	0.0997	0.0613	76%	1.0378	0.0498	0.0625	0.0439	82%
PILS-B	1.0207	0.1095	0.1115	0.0876	87%	1.0077	0.0732	0.0736	0.0624	90%	1.0019	0.0501	0.0502	0.0445	91%
Model C: $h(x)=\{(5/3)x+3/2\}\mathbf{1}\{x<0\}+\{-(4/3)x+3/2\}\mathbf{1}\{x\geq 0\}$															
OLS*	1.0014	0.0383	0.0383	0.0389	96%	1.0018	0.0274	0.0274	0.0276	95%	1.0006	0.0200	0.0200	0.0195	95%
OLS-S	1.2176	0.0924	0.2364	–	–	1.2233	0.0665	0.2330	–	–	1.2231	0.0466	0.2279	–	–
IV-S: $\rho=0.4$	0.9943	0.2260	0.2261	0.2285	96%	0.9984	0.1548	0.1548	0.1622	95%	1.0022	0.1198	0.1198	0.1144	94%
$\rho=0.1$	2.7043	52.7949	52.8224	510.0343	99%	1.0394	0.8935	0.8944	0.8712	99%	0.9791	0.4982	0.4987	0.4988	97%
MSOLS	1.0638	0.0868	0.1077	–	–	1.0566	0.0595	0.0821	–	–	1.0502	0.0417	0.0653	–	–
MSII	0.9784	0.1134	0.1155	–	–	0.9861	0.0730	0.0743	–	–	0.9880	0.0493	0.0507	–	–
MSII-FM	0.9697	0.1132	0.1172	0.0793	84%	0.9842	0.0723	0.0740	0.0520	85%	0.9871	0.0493	0.0509	0.0355	84%
PARA	1.0017	0.1446	0.1446	0.1430	95%	1.0025	0.0996	0.0997	0.1011	96%	0.9974	0.0711	0.0712	0.0716	95%
PILS-E	1.0222	0.0976	0.1001	0.0831	90%	1.0088	0.0632	0.0638	0.0593	94%	1.0022	0.0430	0.0431	0.0427	95%
PILS-B	0.9990	0.0964	0.0964	0.0916	93%	0.9969	0.0632	0.0633	0.0641	95%	0.9936	0.0434	0.0439	0.0450	95%
Model D: $h(x)=x+(5/\tau)\phi(x/\tau)-(5/2)\{\Phi(2/\tau)-1/2\},\ \tau=3/4$															
OLS*	1.0010	0.0446	0.0446	0.0451	95%	1.0023	0.0319	0.0319	0.0320	95%	1.0010	0.0232	0.0232	0.0226	94%
OLS-S	2.3238	0.1118	1.3285	–	–	2.3351	0.0823	1.3376	–	–	2.3345	0.0572	1.3357	–	–
IV-S: $\rho=0.4$	0.9853	0.2838	0.2842	0.2901	96%	0.9924	0.1987	0.1988	0.2057	96%	1.0006	0.1465	0.1465	0.1448	95%
$\rho=0.1$	−1.0251	44.0685	44.1150	699.8918	97%	0.8481	1.2821	1.2911	1.2489	96%	0.9205	0.6967	0.7012	0.6817	95%
MSOLS	1.2823	0.1033	0.3006	–	–	1.2417	0.0694	0.2514	–	–	1.2192	0.0506	0.2250	–	–
MSII	0.8056	0.1766	0.2626	–	–	0.8926	0.1002	0.1469	–	–	0.9362	0.0644	0.0907	–	–
MSII-FM	0.7580	0.1868	0.3057	0.1704	76%	0.8752	0.1022	0.1613	0.0919	74%	0.9280	0.0647	0.0968	0.0564	74%
PARA	1.0172	0.3210	0.3215	0.3149	94%	1.0170	0.2186	0.2193	0.2194	95%	1.0073	0.1527	0.1529	0.1552	95%
PILS-E	1.0768	0.1194	0.1420	0.1182	89%	1.0257	0.0748	0.0791	0.0852	96%	1.0064	0.0519	0.0523	0.0617	98%
PILS-B	0.9631	0.1178	0.1234	0.1382	97%	0.9628	0.0769	0.0854	0.0956	97%	0.9650	0.0528	0.0633	0.0663	97%

Notes: $Mean$ = simulation average of the parameter estimate; SD = simulation standard deviation of the parameter estimate; $RMSE$ = root mean-squared error of the parameter estimate; $\overline{SE}$ = simulation average of the standard error; and CR = coverage rate for the nominal 95% confidence interval.

However, the reduced forms of regressors are different. These are

$$\begin{aligned} X_1 &= \frac{1}{4} X_{3I_C} X_{3E_C}^2 X_{3E_D} + \eta_1, \\ X_2 &= \frac{1}{2} X_{3I_C} X_{3E_C} + \frac{1}{4}\sqrt{25 - X_{3I_C}^2} \sin\left(\frac{\pi}{4} X_{3E_C}\right) X_{3E_D} + \eta_2, \text{ and} \\ X_{3I_C} &= \frac{\sqrt{3}}{2}\rho_1 Z + \frac{3\sqrt{5}}{8}\rho_2 \left(Z^2 - \frac{4}{3}\right) + \frac{\sqrt{3\left(1-\rho_1^2-\rho_2^2\right)}}{2}\eta_2, \end{aligned}$$

where $\rho_1 := corr\left(X_{3I_C}, Z\right)$, $\rho_2 := corr\left(X_{3I_C}, Z^2\right)$, and $\rho_1, \rho_2 \in \{0.1, 0.4\}$. It is worth remarking that $\rho_1 = \rho_2 = 0.4$ roughly mimics the sample correlation coefficients between the endogenous regressor and its IVs in the empirical section of the paper, i.e., $\widehat{corr}\left(educ, fatheduc\right) = 0.4336$ and $\widehat{corr}\left(educ, fatheduc^2\right) = 0.4204$.

The following distributional assumptions also apply.

$$\begin{aligned} &(u, \eta_1)' \sim N\left(0_{2\times 1}, I_2\right). \\ &X_{3E_C}, Z, \eta_2 \overset{iid}{\sim} U\left[-2, 2\right]. \\ &X_{3E_D} \sim Bernoulli\left(1/2\right) - 1/2 = \begin{cases} -1/2 & \text{w.p. } 1/2 \\ 1/2 & \text{w.p. } 1/2 \end{cases}. \\ &u \perp\!\!\!\perp \left(X_{3E_C}, X_{3E_D}, Z, \eta_2\right) \Rightarrow u \perp\!\!\!\perp X_3 = \left(X_{3I_C}, X_{3E_C}, X_{3E_D}\right). \end{aligned}$$

Sample sizes are $n \in \{500, 1000, 2000\}$ and $m = n/2$. The number of replications is 1000. The NNM is based on the Mahalanobis distance, and the number of matches is set equal to $K = 1$ (single match).

This design is largely inspired by the study on nonlinear instruments of Dieterle and Snell (2016). Moreover, the specification of X_{3I_C} ensures compactness of $\operatorname{supp}\left(X_{3I_C}\right)$ and unity of $var\left(X_{3I_C}\right)$ regardless of (ρ_1, ρ_2).

We compare the following estimators of β_3: (i) the infeasible OLS estimator for (9) using $\mathcal{S}^*$ [OLS*]; (ii) the inconsistent OLS estimator for (10) using $\mathcal{S}_1$ only [OLS-S]; (iii) the IV estimator for (10) using $\mathcal{S}_1$ only, where Z is used as an instrument for X_{3I_C} [IV1-S]; (iv) the IV estimator for (10) using $\mathcal{S}_1$ only, where Z^2 is used as an instrument for X_{3I_C} [IV2-S]; (v) the two-step GMM estimator for (10) using $\mathcal{S}_1$ only, where (Z, Z^2) are used as instruments for X_{3I_C}[GMM-S], along with the initial 2SLS estimator [2SLS-S]; (vi) he inconsistent MSOLS estimator for (9) using $\mathcal{S}$ [MSOLS]; (vii) the MSII-FM estimator for (9) using $\mathcal{S}$ and a second-order polynomial [MSII-FM], along with the first-step MSII estimator [MSII]; (viii) the PARA estimator for (9) using $\mathcal{S}_1$ and $\mathcal{S}_2$ [PARA]; (ix) the PILS estimator for (9) using $\mathcal{S}_1$, $\mathcal{S}_2$ and the Epanechnikov kernel [PILS-E]; and (x) the PILS estimator for (9) using $\mathcal{S}_1$, $\mathcal{S}_2$ and the beta kernel [PILS-B]. Implementation methods for two PILS estimators are the same as in Simulation Stydy I.

Note that this design yields that $E(u) = E(X_1 u) = E(X_2 u) = E(X_{3I_C} u) = 0$, which implies that OLS* is consistent. Moreover, $E(v) = E(X_1 v) = 0$, $E(X_{3I_C} v) \neq 0$, and $E(v|Z) = 0$, which leads to consistency of IV1-S, IV2-S, 2SLS-S, and GMM-S.

TABLES 4-5 HERE

Tables 4 and 5 present the results where X_1 enters and is excluded from the regression, respectively. Observe that PARA performs poorly even in the absence of X_1 (= when PARA becomes consistent in theory). Because the reduced form of X_2 is highly nonlinear, the first-step linear approximation is likely to be imprecise (or may be close to lack of identification in a linear form). As a result, PARA tends to have huge bias and variability as is the case with weak instruments. PILS performs exceptionally well.

2.3 Simulation Study III

This simulation study, which deals with yet another case in which X_1 is non-empty, was adopted in the previous version of our paper. The data are generated as follows:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, \tag{13}$$

Table 4: Simulation Study II: Estimates of β_3 in Presence of X_1

	$(n,m)=(500,250)$					$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
Estimator	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*
							$(\rho_1,\rho_2)=(0.1,0.1)$								
OLS*	1.0008	0.0815	0.0815	0.0818	95%	0.9982	0.0574	0.0575	0.0581	95%	1.0016	0.0407	0.0407	0.0411	95%
OLS-S	2.1438	0.0610	1.1454	–	–	2.1439	0.0434	1.1447	–	–	2.1444	0.0301	1.1448	–	–
IV1-S	0.7353	2.7134	2.7263	3.4236	93%	0.8340	0.9000	0.9152	0.8013	94%	0.9522	0.4315	0.4341	0.4223	94%
IV2-S	1.2998	16.3064	16.3092	76.7818	93%	−1.7724	82.9032	82.9496	1618.0879	94%	0.9209	0.4368	0.4440	0.4325	96%
2SLS-S	1.0034	0.7238	0.7238	0.6743	92%	1.0005	0.4134	0.4134	0.4017	93%	0.9990	0.2766	0.2766	0.2745	95%
GMM-S	1.0008	0.7227	0.7227	0.6735	92%	1.0016	0.4141	0.4141	0.4015	92%	0.9991	0.2772	0.2772	0.2745	94%
MSOLS	1.1141	0.0865	0.1432	–	–	1.0809	0.0598	0.1006	–	–	1.0677	0.0419	0.0796	–	–
MSII	0.8366	0.1402	0.2153	–	–	0.9141	0.0745	0.1137	–	–	0.9539	0.0481	0.0666	–	–
MSII-FM	0.7409	0.1515	0.3001	0.1665	74%	0.8709	0.0764	0.1500	0.0875	74%	0.9332	0.0488	0.0827	0.0541	81%
PARA	2.2743	4.2268	4.4147	–	–	2.1140	4.5005	4.6364	–	–	2.2458	2.2119	2.5387	–	–
PILS-E	1.0631	0.1157	0.1318	0.0998	86%	1.0039	0.0723	0.0724	0.0677	94%	0.9891	0.0472	0.0485	0.0461	93%
PILS-B	0.9091	0.1016	0.1363	0.0960	84%	0.9234	0.0674	0.1020	0.0658	78%	0.9446	0.0450	0.0713	0.0454	76%
							$(\rho_1,\rho_2)=(0.1,0.4)$								
OLS*	1.0011	0.0717	0.0717	0.0700	94%	0.9976	0.0505	0.0506	0.0497	95%	1.0008	0.0349	0.0349	0.0352	95%
OLS-S	2.0528	0.0645	1.0548	–	–	2.0520	0.0458	1.0530	–	–	2.0525	0.0318	1.0530	–	–
IV1-S	0.7783	3.6286	3.6354	5.0578	94%	0.7066	4.2328	4.2430	4.4870	95%	0.9523	0.4421	0.4447	0.4302	95%
IV2-S	0.9807	0.1978	0.1987	0.1957	96%	0.9959	0.1344	0.1344	0.1358	95%	0.9937	0.0922	0.0924	0.0960	96%
2SLS-S	0.9977	0.1848	0.1848	0.1877	96%	1.0023	0.1297	0.1298	0.1313	94%	0.9975	0.0906	0.0906	0.0931	95%
GMM-S	0.9958	0.1853	0.1853	0.1873	95%	1.0019	0.1302	0.1302	0.1312	94%	0.9969	0.0905	0.0906	0.0930	95%
MSOLS	1.3346	0.0862	0.3456	–	–	1.3132	0.0615	0.3192	–	–	1.3041	0.0419	0.3070	–	–
MSII	0.8319	0.1765	0.2438	–	–	0.9107	0.1068	0.1392	–	–	0.9547	0.0630	0.0776	–	–
MSII-FM	0.7357	0.1947	0.3283	0.2133	90%	0.8670	0.1105	0.1729	0.1179	87%	0.9338	0.0641	0.0921	0.0730	92%
PARA	2.0323	1.9937	2.2451	–	–	2.1037	1.8392	2.1449	–	–	2.1279	2.2356	2.5040	–	–
PILS-E	1.2343	0.1173	0.2620	0.1002	37%	1.1427	0.0818	0.1645	0.0701	49%	1.0864	0.0514	0.1005	0.0485	55%
PILS-B	0.9637	0.1062	0.1122	0.0975	92%	0.9628	0.0719	0.0809	0.0676	91%	0.9707	0.0476	0.0559	0.0471	90%
							$(\rho_1,\rho_2)=(0.4,0.1)$								
OLS*	1.0022	0.0703	0.0703	0.0702	95%	0.9983	0.0497	0.0497	0.0497	96%	1.0011	0.0357	0.0358	0.0352	95%
OLS-S	2.0541	0.0635	1.0560	–	–	2.0520	0.0463	1.0530	–	–	2.0532	0.0321	1.0537	–	–
IV1-S	0.9882	0.1984	0.1987	0.1950	95%	0.9944	0.1321	0.1322	0.1356	95%	1.0006	0.0950	0.0950	0.0954	94%
IV2-S	0.5743	4.6412	4.6607	7.9092	95%	0.2368	19.9461	19.9607	131.0213	94%	0.9208	0.4555	0.4623	0.4409	96%
2SLS-S	1.0021	0.1881	0.1882	0.1871	95%	1.0013	0.1284	0.1284	0.1312	96%	1.0027	0.0928	0.0928	0.0926	94%
GMM-S	1.0014	0.1882	0.1882	0.1866	95%	1.0004	0.1286	0.1286	0.1310	96%	1.0022	0.0931	0.0931	0.0925	94%
MSOLS	1.3384	0.0841	0.3486	–	–	1.3144	0.0605	0.3201	–	–	1.3056	0.0428	0.3086	–	–
MSII	0.8372	0.1743	0.2385	–	–	0.9154	0.1027	0.1331	–	–	0.9540	0.0649	0.0795	–	–
MSII-FM	0.7416	0.1920	0.3220	0.2118	91%	0.8729	0.1071	0.1662	0.1160	90%	0.9336	0.0660	0.0936	0.0731	91%
PARA	2.0279	1.9303	2.1869	–	–	2.0050	1.7940	2.0564	–	–	2.0949	1.7793	2.0892	–	–
PILS-E	1.2269	0.1149	0.2543	0.0993	39%	1.1361	0.0784	0.1571	0.0695	51%	1.0775	0.0523	0.0935	0.0478	63%
PILS-B	0.9653	0.1048	0.1104	0.0974	93%	0.9633	0.0711	0.0801	0.0675	91%	0.9717	0.0482	0.0559	0.0470	90%
							$(\rho_1,\rho_2)=(0.4,0.4)$								
OLS*	1.0022	0.0635	0.0636	0.0623	94%	0.9977	0.0450	0.0451	0.0441	94%	1.0007	0.0316	0.0316	0.0313	95%
OLS-S	1.9547	0.0675	0.9571	–	–	1.9521	0.0489	0.9533	–	–	1.9529	0.0341	0.9535	–	–
IV1-S	0.9884	0.2023	0.2026	0.1965	95%	0.9954	0.1327	0.1328	0.1364	96%	1.0002	0.0955	0.0955	0.0960	94%
IV2-S	0.9817	0.1985	0.1993	0.1970	96%	0.9970	0.1352	0.1352	0.1366	94%	0.9934	0.0935	0.0937	0.0965	96%
2SLS-S	0.9992	0.1383	0.1383	0.1391	95%	1.0020	0.0968	0.0968	0.0982	96%	0.9995	0.0695	0.0695	0.0695	95%
GMM-S	0.9995	0.1386	0.1386	0.1390	95%	1.0024	0.0969	0.0969	0.0982	96%	0.9997	0.0696	0.0696	0.0695	95%
MSOLS	1.4173	0.0901	0.4269	–	–	1.3986	0.0644	0.4038	–	–	1.3913	0.0455	0.3939	–	–
MSII	0.8289	0.2343	0.2902	–	–	0.9177	0.1310	0.1547	–	–	0.9540	0.0825	0.0945	–	–
MSII-FM	0.7530	0.2564	0.3560	0.2519	97%	0.8865	0.1350	0.1764	0.1328	93%	0.9392	0.0838	0.1035	0.0841	94%
PARA	1.9433	1.5188	1.7879	–	–	2.0057	1.7421	2.0116	–	–	1.9520	1.7781	2.0169	–	–
PILS-E	1.2932	0.1102	0.3132	0.0958	18%	1.2054	0.0811	0.2208	0.0679	18%	1.1398	0.0539	0.1498	0.0475	19%
PILS-B	1.0082	0.1068	0.1072	0.0947	91%	0.9949	0.0740	0.0742	0.0658	93%	0.9934	0.0496	0.0501	0.0460	92%

Table 5: Simulation Study II: Estimates of β_3 in Absence of X_1

	$(n,m)=(500,250)$					$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
Estimator	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*
							$(\rho_1,\rho_2)=(0.1,0.1)$								
OLS*	0.9997	0.0795	0.0795	0.0821	95%	1.0002	0.0568	0.0568	0.0581	95%	0.9983	0.0415	0.0416	0.0411	94%
OLS-S	2.1416	0.0607	1.1432	–	–	2.1426	0.0418	1.1434	–	–	2.1428	0.0294	1.1432	–	–
IV1-S	0.8158	2.9035	2.9094	3.4401	94%	0.9130	2.1501	2.1519	2.1972	94%	0.9276	0.4427	0.4485	0.4330	95%
IV2-S	0.5557	5.9991	6.0155	12.6844	93%	0.9142	1.6621	1.6643	1.2972	94%	0.9109	0.4599	0.4684	0.4431	96%
2SLS-S	1.0007	0.7826	0.7826	0.6656	90%	1.0104	0.3961	0.3963	0.3990	94%	0.9931	0.2607	0.2608	0.2765	95%
GMM-S	0.9979	0.7879	0.7879	0.6649	90%	1.0102	0.3963	0.3964	0.3988	94%	0.9931	0.2608	0.2609	0.2764	95%
MSOLS	1.1122	0.0850	0.1407	–	–	1.0823	0.0616	0.1028	–	–	1.0645	0.0441	0.0781	–	–
MSII	0.8366	0.1311	0.2095	–	–	0.9127	0.0779	0.1170	–	–	0.9499	0.0507	0.0713	–	–
MSII-FM	0.7417	0.1425	0.2950	0.1639	72%	0.8698	0.0804	0.1530	0.0888	76%	0.9294	0.0514	0.0873	0.0544	79%
PARA	2.1833	1.9409	2.2731	3.2079	83%	2.2256	2.1668	2.4894	3.9499	85%	2.1041	2.4597	2.6962	4.7421	86%
PILS-E	1.0572	0.1126	0.1263	0.1001	88%	1.0061	0.0748	0.0751	0.0677	92%	0.9853	0.0484	0.0506	0.0460	92%
PILS-B	0.9063	0.0997	0.1368	0.0963	83%	0.9247	0.0684	0.1017	0.0657	79%	0.9417	0.0463	0.0745	0.0454	75%
							$(\rho_1,\rho_2)=(0.1,0.4)$								
OLS*	0.9998	0.0682	0.0682	0.0702	96%	1.0009	0.0475	0.0475	0.0497	97%	0.9988	0.0356	0.0357	0.0352	95%
OLS-S	2.0507	0.0645	1.0527	–	–	2.0517	0.0447	1.0527	–	–	2.0517	0.0315	1.0522	–	–
IV1-S	0.4841	4.9393	4.9662	12.0519	96%	0.8377	0.9188	0.9330	0.8795	94%	0.9374	0.4313	0.4358	0.4351	96%
IV2-S	0.9935	0.1987	0.1988	0.1936	96%	0.9964	0.1346	0.1346	0.1361	96%	0.9912	0.0974	0.0978	0.0963	95%
2SLS-S	1.0053	0.1919	0.1920	0.1863	95%	1.0029	0.1305	0.1305	0.1317	95%	0.9947	0.0926	0.0928	0.0934	95%
GMM-S	1.0038	0.1922	0.1923	0.1858	95%	1.0020	0.1304	0.1304	0.1315	95%	0.9944	0.0927	0.0929	0.0933	95%
MSOLS	1.3338	0.0902	0.3458	–	–	1.3151	0.0634	0.3215	–	–	1.3009	0.0442	0.3041	–	–
MSII	0.8319	0.1807	0.2468	–	–	0.9168	0.1055	0.1343	–	–	0.9490	0.0651	0.0827	–	–
MSII-FM	0.7362	0.1957	0.3285	0.2146	91%	0.8727	0.1106	0.1686	0.1165	88%	0.9279	0.0667	0.0982	0.0734	89%
PARA	2.0047	2.3243	2.5322	4.8806	82%	2.0878	5.7381	5.8403	24.6909	84%	2.0599	2.2652	2.5009	4.4293	83%
PILS-E	1.2355	0.1208	0.2647	0.1007	39%	1.1442	0.0811	0.1654	0.0702	46%	1.0837	0.0539	0.0996	0.0483	57%
PILS-B	0.9601	0.1053	0.1126	0.0977	92%	0.9643	0.0731	0.0813	0.0675	90%	0.9683	0.0496	0.0588	0.0470	88%
							$(\rho_1,\rho_2)=(0.4,0.1)$								
OLS*	0.9994	0.0696	0.0696	0.0702	95%	1.0001	0.0482	0.0482	0.0497	95%	0.9989	0.0350	0.0351	0.0352	95%
OLS-S	2.0508	0.0652	1.0528	–	–	2.0516	0.0443	1.0525	–	–	2.0513	0.0314	1.0518	–	–
IV1-S	0.9860	0.1939	0.1944	0.1941	96%	0.9949	0.1340	0.1341	0.1357	95%	0.9951	0.0924	0.0925	0.0956	96%
IV2-S	−2.7750	101.6185	101.6886	3079.0631	94%	0.8782	1.2733	1.2791	1.0141	95%	0.9189	0.4662	0.4732	0.4485	96%
2SLS-S	1.0020	0.1877	0.1877	0.1864	95%	1.0019	0.1293	0.1293	0.1313	94%	0.9975	0.0883	0.0884	0.0929	96%
GMM-S	0.9994	0.1887	0.1887	0.1860	95%	1.0008	0.1295	0.1295	0.1311	94%	0.9972	0.0884	0.0885	0.0928	96%
MSOLS	1.3379	0.0877	0.3491	–	–	1.3172	0.0612	0.3230	–	–	1.3039	0.0432	0.3070	–	–
MSII	0.8345	0.1870	0.2497	–	–	0.9168	0.1011	0.1309	–	–	0.9514	0.0652	0.0813	–	–
MSII-FM	0.7371	0.2090	0.3358	0.2168	91%	0.8737	0.1049	0.1642	0.1162	88%	0.9310	0.0663	0.0957	0.0731	90%
PARA	2.1785	1.8965	2.2328	3.2866	81%	2.1775	1.9522	2.2799	3.4303	84%	2.0936	2.8626	3.0644	5.8310	84%
PILS-E	1.2330	0.1205	0.2623	0.1001	38%	1.1369	0.0809	0.1590	0.0694	48%	1.0761	0.0527	0.0926	0.0477	64%
PILS-B	0.9632	0.1070	0.1132	0.0975	92%	0.9645	0.0739	0.0820	0.0674	91%	0.9700	0.0496	0.0580	0.0469	89%
							$(\rho_1,\rho_2)=(0.4,0.4)$								
OLS*	0.9996	0.0617	0.0617	0.0623	95%	1.0006	0.0422	0.0422	0.0442	97%	0.9992	0.0310	0.0310	0.0313	96%
OLS-S	1.9511	0.0689	0.9536	–	–	1.9521	0.0477	0.9533	–	–	1.9518	0.0336	0.9524	–	–
IV1-S	0.9862	0.1945	0.1949	0.1950	96%	0.9951	0.1347	0.1348	0.1364	95%	0.9960	0.0922	0.0923	0.0961	96%
IV2-S	0.9940	0.1987	0.1988	0.1945	95%	0.9968	0.1358	0.1358	0.1369	95%	0.9922	0.0973	0.0976	0.0967	95%
2SLS-S	1.0024	0.1415	0.1415	0.1386	94%	1.0018	0.0976	0.0976	0.0984	95%	0.9972	0.0661	0.0661	0.0697	96%
GMM-S	1.0027	0.1418	0.1418	0.1385	94%	1.0018	0.0976	0.0976	0.0983	95%	0.9973	0.0661	0.0662	0.0697	96%
MSOLS	1.4161	0.0932	0.4264	–	–	1.4026	0.0647	0.4078	–	–	1.3924	0.0458	0.3950	–	–
MSII	0.8341	0.2340	0.2868	–	–	0.9204	0.1287	0.1514	–	–	0.9563	0.0831	0.0939	–	–
MSII-FM	0.7591	0.2556	0.3512	0.2461	96%	0.8875	0.1326	0.1739	0.1335	94%	0.9412	0.0843	0.1028	0.0839	93%
PARA	2.0331	1.5480	1.8611	2.9925	80%	1.9910	2.2472	2.4560	4.8988	83%	1.9656	2.3171	2.5103	4.0836	83%
PILS-E	1.2947	0.1130	0.3157	0.0960	19%	1.2112	0.0780	0.2251	0.0681	17%	1.1396	0.0558	0.1504	0.0474	21%
PILS-B	1.0038	0.1112	0.1113	0.0945	90%	0.9972	0.0733	0.0734	0.0657	93%	0.9925	0.0499	0.0505	0.0459	92%

where true parameter values are $\beta_0 = \beta_1 = \beta_2 = 1$, and β_1 is the parameter of interest. There are two common variables, namely, a continuous one X_{3E_C} and a discrete one X_{3E_D}, and both are excluded from the regression (13). The regressors X_1 and X_2 have the reduced forms

$$X_1 = 1 + X_{3E_C}^2 + X_{3E_D} + (\pi Z + \xi) \text{ and } X_2 = \{h(X_{3E_C}) + X_{3E_D}\} + \eta_2,$$

where Z is the instrument for X_1 and the parameter π, whose values will be presented shortly, controls strength of the instrument. In particular, Z is applied to the short regression after omitting X_2 from (13), i.e.,

$$Y = \beta_0 + \beta_1 X_1 + (u + \beta_2 X_2) =: \beta_0 + \beta_1 X_1 + v. \tag{14}$$

The function $h(\cdot)$ is specified as one of the followings:

$$h(x) = \begin{cases} x & \text{[Model A]} \\ \tanh(x) = \{\exp(x) - \exp(-x)\} / \{\exp(x) + \exp(-x)\} & \text{[Model B]} \\ (3/4)\sqrt{x+2} - 1 & \text{[Model C]} \\ x^3/2 - x^2/2 - 2x - 2/3 & \text{[Model D]} \\ |x| - 1 & \text{[Model E]} \\ x + (5/\tau)\phi(x/\tau) - (5/2)\{\Phi(2/\tau) - 1/2\}, \ \tau = 3/4 & \text{[Model F]} \end{cases}.$$

Observe that Models A and F are exactly the same as Models A and D in Simulation Stydy I, respectively. The following distributional assumptions also apply.

$$\begin{aligned} &(u, \eta_2, \xi)' \sim N(0_{3\times 1}, I_3). \\ &Z, X_{3E_C} \overset{iid}{\sim} U[-2, 2]. \\ &X_{3E_D} \sim \mathit{Bernoulli}(1/2) - 1/2 = \begin{cases} -1/2 & \text{w.p. } 1/2 \\ 1/2 & \text{w.p. } 1/2 \end{cases}. \\ &u \perp\!\!\!\perp X_3 = (X_{3E_C}, X_{3E_D}). \end{aligned}$$

The values of π are $\pi \in \{1.00, 0.70, 0.15\}$. It follows from the distributional assumptions that

$$\mathit{corr}(X_1, Z) = \sqrt{\frac{240\pi^2}{240\pi^2 + 481}} \approx \begin{cases} 0.58 & \text{for } \pi = 1.00 \\ 0.44 & \text{for } \pi = 0.70 \\ 0.11 & \text{for } \pi = 0.15 \end{cases}.$$

In short, $\pi \in \{1.00, 0.70\}$ and $\pi = 0.15$ correspond to strong and weak IV cases, and the correlation for $\pi = 0.70$ is roughly similar to the sample correlation between the endogenous variable (*educ*) and its instrument (*fatheduc*) in our empirical example in Section 4 of the paper.

The above procedure provides us with two observable samples

$$\mathcal{S}_1 = \{(Y_i, X_{1i}, X_{3E_C i}, X_{3E_D i}, Z_i)\}_{i=1}^{n} \text{ and } \mathcal{S}_2 = \{(X_{2j}, X_{3E_C j}, X_{3E_D j})\}_{j=1}^{m}.$$

The complete sample

$$\mathcal{S}^* = \{(Y_i, X_{1i}, X_{2i}, X_{3E_C i}, X_{3E_D i}, Z_i)\}_{i=1}^{n}$$

is the sample that would not be observed in practice. Finally, the matched sample

$$\mathcal{S} = \left\{\left(Y_i, X_{1i}, X_{2j_1(i)}, \ldots, X_{2j_K(i)}, X_{3E_C i}, X_{3E_D i}, Z_i\right)\right\}_{i=1}^{n}$$

is constructed via the NNM with respect to $X_3 = (X_{3E_C}, X_{3E_D})$ as in Hirukawa and Prokhorov (2018). The NNM is based on the Mahalanobis metric and the number of matches $K \in \{1, 2, 4, 8, 16, 32, 64, 128\}$. As regards the sample sizes, we choose $n \in \{500, 1000\}$ and put $m = n$ for simplicity. For each combination of the functional form of $h(\cdot)$ and the value of π, 1000 Monte Carlo samples are drawn.

We compare the following estimators of β_1: (i) the infeasible OLS estimator for (13) using $\mathcal{S}^*$ [OLS*]; (ii) the inconsistent OLS estimator for (14) using $\mathcal{S}_1$ only [OLS-S]; (iii) the IV estimator for (14) using $\mathcal{S}_1$ only [IV-S]; (iv) the inconsistent MSOLS estimator for (13) using $\mathcal{S}$ [MSOLS]; (v) the MSII estimator for (13) using $\mathcal{S}$ [MSII]; (vi) the PARA estimator for (13) using $\mathcal{S}_1$ and $\mathcal{S}_2$ [PARA]; (vii) the PILS estimator for (13) using $\mathcal{S}_1$, $\mathcal{S}_2$ and the Epanechnikov kernel [PILS-E]; and (viii) the PILS estimator for (13) using $\mathcal{S}_1$, $\mathcal{S}_2$ and the beta kernel [PILS-B]. Implementation methods for two PILS estimators are the same as in Simulation Stydy I.

As in Simulation Studies I and II, OLS* is consistent, because $E(u) = E(X_1 u) = E(X_2 u) = 0$. Moreover, $E(v) = 0$, $E(X_1 v) \neq 0$, $corr(X_1, Z) \neq 0$ and $E(Zv) = 0$, which implies inconsistency of OLS-S and consistency of IV-S. Furthermore, as indicated in Proposition S2, PARA becomes consistent for an odd $h(\cdot)$, i.e., for Models A and B. Because of inconsistency we do not compute $\overline{SE}$ or CR of OLS-S and MSOLS. Likewise, we present $\overline{SE}$ and CR of PARA only when it becomes consistent, i.e., for Models A and B.

TABLE 6 HERE

Table 6: Simulation Study III: Estimates of β_1

$n\,(=m)$	π	Estimator	OLS*	OLS-S	IV-S	MSOLS	MSII	PARA	PILS-E	PILS-B
					Model A: $h(x)=x$					
500	1.00	$Mean$	0.9984	1.0611	0.9956	1.0227	0.9975	0.9989	1.0069	1.0050
		SD	0.0224	0.0425	0.0737	0.0373	0.0427	0.0318	0.0347	0.0345
		$RMSE$	0.0225	0.0744	0.0738	0.0436	0.0428	0.0318	0.0354	0.0349
		$\overline{SE}$	0.0223	–	0.0738	–	0.0395	0.0321	0.0389	0.0393
		CR	95%	–	95%	–	93%	95%	97%	97%
	0.70	$Mean$	0.9985	1.0740	0.9934	1.0278	0.9975	0.9992	1.0088	1.0065
		SD	0.0245	0.0470	0.1063	0.0413	0.0476	0.0351	0.0389	0.0386
		$RMSE$	0.0246	0.0876	0.1065	0.0498	0.0477	0.0351	0.0399	0.0392
		$\overline{SE}$	0.0245	–	0.1062	–	0.0435	0.0353	0.0427	0.0432
		CR	95%	–	95%	–	92%	96%	97%	98%
	0.15	$Mean$	0.9992	1.0920	0.8116	1.0353	0.9979	1.0000	1.0118	1.0091
		SD	0.0270	0.0529	2.7255	0.0468	0.0542	0.0394	0.0445	0.0443
		$RMSE$	0.0271	0.1061	2.7320	0.0586	0.0543	0.0394	0.0461	0.0452
		$\overline{SE}$	0.0272	0.0536	5.5917	0.0476	0.0483	0.0394	0.0474	0.0480
		CR	95%	–	99%	–	92%	95%	97%	97%
1000	1.00	$Mean$	1.0003	1.0636	1.0005	1.0262	1.0020	1.0008	1.0068	1.0055
		SD	0.0159	0.0313	0.0498	0.0273	0.0307	0.0227	0.0247	0.0246
		$RMSE$	0.0159	0.0709	0.0498	0.0378	0.0308	0.0227	0.0256	0.0252
		$\overline{SE}$	0.0158	–	0.0519	–	0.0278	0.0227	0.0275	0.0277
		CR	95%	–	96%	–	93%	95%	96%	96%
	0.70	$Mean$	1.0004	1.0765	1.0006	1.0314	1.0022	1.0009	1.0081	1.0065
		SD	0.0175	0.0346	0.0712	0.0302	0.0343	0.0250	0.0276	0.0275
		$RMSE$	0.0175	0.0840	0.0712	0.0436	0.0344	0.0250	0.0288	0.0283
		$\overline{SE}$	0.0174	–	0.0744	–	0.0306	0.0250	0.0302	0.0305
		CR	95%	–	96%	–	92%	96%	96%	96%
	0.15	$Mean$	1.0006	1.0940	1.0041	1.0383	1.0021	1.0008	1.0096	1.0077
		SD	0.0193	0.0382	0.4137	0.0336	0.0387	0.0278	0.0314	0.0312
		$RMSE$	0.0193	0.1014	0.4137	0.0509	0.0387	0.0279	0.0328	0.0322
		$\overline{SE}$	0.0193	–	0.4534	–	0.0340	0.0279	0.0336	0.0338
		CR	95%	–	98%	–	91%	96%	95%	96%
					Model B: $h(x)=\tanh(x)=\{\exp(x)-\exp(-x)\}\,/\,\{\exp(x)+\exp(-x)\}$					
500	1.00	$Mean$	0.9984	1.0614	0.9955	1.0342	0.9960	0.9990	1.0069	1.0051
		SD	0.0224	0.0361	0.0643	0.0351	0.0443	0.0321	0.0347	0.0345
		$RMSE$	0.0225	0.0712	0.0644	0.0490	0.0445	0.0321	0.0354	0.0349
		$\overline{SE}$	0.0224	–	0.0649	–	0.0415	0.0324	0.0391	0.0395
		CR	95%	–	95%	–	94%	96%	97%	97%
	0.70	$Mean$	0.9985	1.0744	0.9933	1.0417	0.9956	0.9993	1.0088	1.0066
		SD	0.0245	0.0397	0.0927	0.0388	0.0497	0.0354	0.0389	0.0387
		$RMSE$	0.0246	0.0843	0.0929	0.0570	0.0499	0.0355	0.0399	0.0393
		$\overline{SE}$	0.0245	–	0.0934	–	0.0459	0.0357	0.0429	0.0434
		CR	95%	–	95%	–	94%	96%	97%	98%
	0.15	$Mean$	0.9991	1.0925	0.8818	1.0525	0.9956	1.0002	1.0119	1.0092
		SD	0.0271	0.0446	1.7635	0.0438	0.0570	0.0399	0.0446	0.0444
		$RMSE$	0.0271	0.1027	1.7675	0.0684	0.0571	0.0399	0.0462	0.0454
		$\overline{SE}$	0.0272	–	3.4571	–	0.0515	0.0398	0.0477	0.0483
		CR	95%	–	99%	–	93%	95%	96%	97%
1000	1.00	$Mean$	1.0003	1.0636	1.0006	1.0373	1.0013	1.0010	1.0067	1.0056
		SD	0.0159	0.0267	0.0440	0.0257	0.0315	0.0229	0.0248	0.0247
		$RMSE$	0.0159	0.0690	0.0440	0.0453	0.0316	0.0229	0.0257	0.0253
		$\overline{SE}$	0.0158	–	0.0456	–	0.0287	0.0229	0.0277	0.0279
		CR	95%	–	96%	–	93%	95%	95%	96%
	0.70	$Mean$	1.0004	1.0766	1.0007	1.0449	1.0014	1.0010	1.0080	1.0066
		SD	0.0175	0.0293	0.0629	0.0283	0.0354	0.0253	0.0278	0.0277
		$RMSE$	0.0175	0.0820	0.0629	0.0531	0.0354	0.0253	0.0289	0.0284
		$\overline{SE}$	0.0174	–	0.0654	–	0.0318	0.0252	0.0304	0.0306
		CR	95%	–	96%	–	92%	96%	95%	96%
	0.15	$Mean$	1.0007	1.0940	1.0037	1.0550	1.0012	1.0010	1.0095	1.0078
		SD	0.0193	0.0321	0.3648	0.0313	0.0401	0.0282	0.0316	0.0315
		$RMSE$	0.0194	0.0993	0.3648	0.0633	0.0401	0.0282	0.0330	0.0324
		$\overline{SE}$	0.0193	–	0.3981	–	0.0356	0.0282	0.0338	0.0341
		CR	95%	–	98%	–	92%	95%	95%	96%

Table 6: *(cont'd)*

$n\,(=m)$	π	Estimator	OLS*	OLS-S	IV-S	MSOLS	MSII	PARA	PILS-E	PILS-B
		Model C: $h\,(x) = (3/4)\,\sqrt{x+2} - 1$								
500	1.00	$Mean$	0.9984	1.0424	0.9958	1.0305	0.9926	0.9792	1.0053	1.0025
		SD	0.0224	0.0335	0.0599	0.0335	0.0615	0.0321	0.0349	0.0347
		$RMSE$	0.0225	0.0541	0.0601	0.0453	0.0619	0.0383	0.0353	0.0348
		$\overline{SE}$	0.0223	–	0.0601	–	0.0523	–	0.0393	0.0399
		CR	94%	–	96%	–	95%	–	97%	98%
	0.70	$Mean$	0.9985	1.0516	0.9938	1.0373	0.9913	0.9752	1.0068	1.0035
		SD	0.0246	0.0370	0.0864	0.0369	0.0704	0.0355	0.0391	0.0389
		$RMSE$	0.0246	0.0634	0.0867	0.0525	0.0710	0.0433	0.0397	0.0391
		$\overline{SE}$	0.0245	–	0.0865	–	0.0587	–	0.0432	0.0439
		CR	95%	–	96%	–	96%	–	97%	98%
	0.15	$Mean$	0.9991	1.0644	0.9603	1.0469	0.9900	0.9701	1.0095	1.0053
		SD	0.0271	0.0416	1.2262	0.0416	0.0811	0.0399	0.0449	0.0448
		$RMSE$	0.0271	0.0767	1.2268	0.0627	0.0817	0.0498	0.0459	0.0451
		$\overline{SE}$	0.0272	–	2.2439	–	0.0665	–	0.0480	0.0488
		CR	95%	–	99%	–	96%	–	96%	97%
1000	1.00	$Mean$	1.0003	1.0444	1.0004	1.0331	1.0001	0.9811	1.0055	1.0036
		SD	0.0159	0.0246	0.0407	0.0244	0.0332	0.0228	0.0248	0.0247
		$RMSE$	0.0159	0.0508	0.0407	0.0411	0.0332	0.0296	0.0254	0.0250
		$\overline{SE}$	0.0158	–	0.0423	–	0.0305	–	0.0278	0.0281
		CR	95%	–	96%	–	94%	–	96%	96%
	0.70	$Mean$	1.0004	1.0534	1.0005	1.0398	0.9998	0.9769	1.0065	1.0042
		SD	0.0175	0.0270	0.0582	0.0268	0.0374	0.0253	0.0278	0.0277
		$RMSE$	0.0175	0.0599	0.0582	0.0480	0.0374	0.0342	0.0285	0.0280
		$\overline{SE}$	0.0174	–	0.0605	–	0.0340	–	0.0306	0.0309
		CR	95%	–	96%	–	93%	–	96%	96%
	0.15	$Mean$	1.0006	1.0656	1.0054	1.0488	0.9993	0.9709	1.0076	1.0049
		SD	0.0193	0.0297	0.3357	0.0296	0.0427	0.0283	0.0317	0.0315
		$RMSE$	0.0193	0.0720	0.3358	0.0571	0.0427	0.0406	0.0326	0.0319
		$\overline{SE}$	0.0193	–	0.3662	–	0.0384	–	0.0340	0.0343
		CR	96%	–	98%	–	93%	–	95%	96%
		Model D: $h\,(x) = x^3/2 - x^2/2 - 2x - 2/3$								
500	1.00	$Mean$	0.9986	0.8841	0.9976	0.9574	1.0001	0.8185	1.0021	0.9957
		SD	0.0227	0.0423	0.0797	0.0382	0.0435	0.0362	0.0353	0.0350
		$RMSE$	0.0228	0.1233	0.0797	0.0572	0.0435	0.1851	0.0353	0.0353
		$\overline{SE}$	0.0225	–	0.0762	–	0.0402	–	0.0396	0.0400
		CR	94%	–	93%	–	93%	–	98%	98%
	0.70	$Mean$	0.9987	0.8608	0.9972	0.9490	1.0007	0.7810	1.0031	0.9953
		SD	0.0249	0.0458	0.1149	0.0422	0.0488	0.0398	0.0397	0.0394
		$RMSE$	0.0250	0.1465	0.1150	0.0662	0.0488	0.2226	0.0398	0.0396
		$\overline{SE}$	0.0247	–	0.1098	–	0.0445	–	0.0436	0.0441
		CR	95%	–	93%	–	92%	–	98%	97%
	0.15	$Mean$	0.9993	0.8297	1.2919	0.9380	1.0020	0.7303	1.0050	0.9953
		SD	0.0276	0.0502	4.1777	0.0477	0.0562	0.0447	0.0458	0.0454
		$RMSE$	0.0276	0.1775	4.1878	0.0782	0.0562	0.2734	0.0461	0.0457
		$\overline{SE}$	0.0274	–	8.3371	–	0.0498	–	0.0486	0.0491
		CR	95%	–	99%	–	92%	–	97%	97%
1000	1.00	$Mean$	1.0002	0.8852	1.0001	0.9602	1.0030	0.8202	1.0036	0.9989
		SD	0.0159	0.0284	0.0526	0.0270	0.0313	0.0254	0.0250	0.0249
		$RMSE$	0.0159	0.1182	0.0526	0.0481	0.0314	0.1816	0.0253	0.0250
		$\overline{SE}$	0.0159	–	0.0536	–	0.0284	–	0.0280	0.0281
		CR	95%	–	95%	–	92%	–	96%	96%
	0.70	$Mean$	1.0003	0.8618	1.0004	0.9517	1.0034	0.7827	1.0042	0.9985
		SD	0.0175	0.0310	0.0752	0.0300	0.0351	0.0281	0.0281	0.0280
		$RMSE$	0.0175	0.1416	0.0752	0.0568	0.0353	0.2191	0.0284	0.0280
		$\overline{SE}$	0.0175	–	0.0768	–	0.0314	–	0.0308	0.0310
		CR	94%	–	95%	–	92%	–	96%	96%
	0.15	$Mean$	1.0005	0.8299	1.0313	0.9399	1.0036	0.7312	1.0050	0.9979
		SD	0.0195	0.0344	0.4877	0.0338	0.0398	0.0318	0.0320	0.0318
		$RMSE$	0.0195	0.1735	0.4887	0.0689	0.0400	0.2706	0.0324	0.0319
		$\overline{SE}$	0.0195	–	0.4672	–	0.0351	–	0.0343	0.0345
		CR	95%	–	98%	–	92%	–	96%	96%

Table 6: *(cont'd)*

$n\,(=m)$	π	Estimator	OLS*	OLS-S	IV-S	MSOLS	MSII	PARA	PILS-E	PILS-B
					Model E: $h\,(x) = \lvert x\rvert - 1$					
500	1.00	*Mean*	0.9981	1.2282	0.9939	1.1647	0.9666	1.1770	1.0116	1.0061
		SD	0.0237	0.0332	0.0620	0.0362	0.1397	0.0334	0.0398	0.0397
		RMSE	0.0238	0.2306	0.0623	0.1686	0.1436	0.1801	0.0414	0.0402
		$\overline{SE}$	0.0239	–	0.0628	–	0.1173	–	0.0482	0.0495
		CR	96%	–	95%	–	97%	–	98%	99%
	0.70	*Mean*	0.9981	1.2753	0.9905	1.2055	0.9308	1.2166	1.0164	1.0091
		SD	0.0264	0.0365	0.0896	0.0405	0.4623	0.0370	0.0467	0.0467
		RMSE	0.0264	0.2777	0.0901	0.2094	0.4675	0.2198	0.0495	0.0476
		$\overline{SE}$	0.0266	–	0.0906	–	0.3786	–	0.0560	0.0576
		CR	96%	–	96%	–	96%	–	97%	98%
	0.15	*Mean*	0.9987	1.3397	0.9013	1.2658	0.9728	1.2726	1.0262	1.0153
		SD	0.0298	0.0408	1.3240	0.0463	1.1735	0.0416	0.0582	0.0586
		RMSE	0.0299	0.3421	1.3276	0.2698	1.1738	0.2758	0.0638	0.0606
		$\overline{SE}$	0.0302	–	1.9154	–	1.9124	–	0.0677	0.0700
		CR	96%	–	98%	–	95%	–	95%	97%
1000	1.00	*Mean*	1.0004	1.2297	0.9997	1.1674	0.9893	1.1785	1.0081	1.0048
		SD	0.0172	0.0241	0.0423	0.0256	0.0656	0.0239	0.0277	0.0278
		RMSE	0.0172	0.2310	0.0423	0.1694	0.0665	0.1801	0.0288	0.0282
		$\overline{SE}$	0.0169	–	0.0441	–	0.0640	–	0.0342	0.0348
		CR	94%	–	96%	–	96%	–	98%	99%
	0.70	*Mean*	1.0006	1.2767	0.9992	1.2080	0.9830	1.2179	1.0107	1.0062
		SD	0.0193	0.0263	0.0605	0.0285	0.0896	0.0264	0.0327	0.0329
		RMSE	0.0193	0.2779	0.0605	0.2099	0.0912	0.2195	0.0344	0.0335
		$\overline{SE}$	0.0189	–	0.0632	–	0.0862	–	0.0399	0.0406
		CR	95%	–	96%	–	96%	–	98%	98%
	0.15	*Mean*	1.0009	1.3404	0.9611	1.2675	0.9662	1.2731	1.0151	1.0085
		SD	0.0217	0.0285	0.3724	0.0316	0.1514	0.0292	0.0411	0.0415
		RMSE	0.0217	0.3416	0.3745	0.2693	0.1551	0.2747	0.0438	0.0424
		$\overline{SE}$	0.0214	–	0.4086	–	0.1379	–	0.0486	0.0497
		CR	95%	–	97%	–	94%	–	96%	97%
					Model F: $h\,(x) = x + (5/\tau)\,\phi\,(x/\tau) - (5/2)\,\{\Phi\,(2/\tau) - 1/2\}\,,\ \tau = 3/4$					
500	1.00	*Mean*	0.9986	0.8115	0.9980	0.9413	1.0001	0.7468	1.0048	1.0018
		SD	0.0231	0.0445	0.0811	0.0388	0.0446	0.0349	0.0353	0.0352
		RMSE	0.0231	0.1937	0.0812	0.0703	0.0446	0.2556	0.0356	0.0353
		$\overline{SE}$	0.0227	–	0.0817	–	0.0412	–	0.0401	0.0407
		CR	94%	–	95%	–	93%	–	98%	98%
	0.70	*Mean*	0.9987	0.7734	0.9978	0.9292	1.0007	0.6948	1.0064	1.0028
		SD	0.0254	0.0489	0.1170	0.0431	0.0503	0.0382	0.0398	0.0397
		RMSE	0.0255	0.2318	0.1170	0.0829	0.0504	0.3076	0.0403	0.0398
		$\overline{SE}$	0.0250	–	0.1176	–	0.0459	–	0.0443	0.0450
		CR	95%	–	96%	–	92%	–	98%	97%
	0.15	*Mean*	0.9994	0.7222	0.9601	0.9130	1.0022	0.6246	1.0093	1.0047
		SD	0.0283	0.0546	2.6560	0.0491	0.0584	0.0426	0.0461	0.0460
		RMSE	0.0283	0.2831	2.6563	0.0999	0.0584	0.3778	0.0470	0.0462
		$\overline{SE}$	0.0279	–	5.2679	–	0.0518	–	0.0496	0.0504
		CR	95%	–	99%	–	92%	–	97%	97%
1000	1.00	*Mean*	1.0000	0.8144	1.0015	0.9447	1.0030	0.7490	1.0057	1.0037
		SD	0.0164	0.0328	0.0565	0.0284	0.0318	0.0246	0.0252	0.0253
		RMSE	0.0164	0.1885	0.0565	0.0621	0.0320	0.2522	0.0259	0.0255
		$\overline{SE}$	0.0161	–	0.0575	–	0.0292	–	0.0283	0.0286
		CR	94%	–	96%	–	92%	–	97%	97%
	0.70	*Mean*	1.0000	0.7763	1.0026	0.9326	1.0034	0.6969	1.0068	1.0044
		SD	0.0181	0.0361	0.0809	0.0316	0.0360	0.0271	0.0285	0.0285
		RMSE	0.0181	0.2266	0.0809	0.0745	0.0362	0.3044	0.0293	0.0288
		$\overline{SE}$	0.0178	–	0.0824	–	0.0324	–	0.0313	0.0316
		CR	94%	–	96%	–	92%	–	97%	97%
	0.15	*Mean*	1.0002	0.7243	1.0641	0.9155	1.0038	0.6254	1.0083	1.0052
		SD	0.0202	0.0400	0.6229	0.0356	0.0414	0.0303	0.0326	0.0326
		RMSE	0.0202	0.2785	0.6262	0.0917	0.0415	0.3758	0.0337	0.0330
		$\overline{SE}$	0.0198	–	0.6042	–	0.0366	–	0.0350	0.0354
		CR	95%	–	98%	–	91%	–	96%	97%

Results are presented in Table 6. For MSII and MSOLS we report only the results for $K = 1$ (single match), because those for $K \geq 2$ are qualitatively similar.

Moreover, since the regressor X_1 depends on π in this design, the results for all estimators, not only for IV-S, vary with the strength of the instrument, unlike the design in Section 3 of the paper. It is indicated that PILS-B again works well for a wide variety of models.

2.4 Simulation Study IV

This study is a variation of what ended up in the paper – it uses the setup of Section 3.1 of the paper, except that X_1 is handled differently. In the original setup, the regressor X_1 is merely orthogonal to the composite error term v. Here, we generate X_1 as $X_1 \overset{iid}{\sim} N(0,1)$, independently of all random variables in the design. Notice that the data generation of all other variables is exactly the same as before. It follows from $X_1 \perp\!\!\!\perp (u, X_2, Z)$ that the conditional moment restriction $E(v|X_1, Z) = 0$ holds for (S). Results for this setup are presented in Table 7.

TABLE 7 HERE

3 More on the Empirical Application

In this section, a wider range of estimators for the return to schooling application including nonlinear IV are presented. Subsequently, we argue that if the PSID ability measure is error-ridden, OLS* is upward-inconsistent suggesting that PILS is likely to be the only estimator that delivers the closest estimate to the true population value. We conclude this section by conducting sensitivity analysis in PILS to the choices of smoothing parameter values.

Table 7: Simulation Study IV: Estimates of β_3 under Independent X_1

	$(n,m)=(1000,500)$					$(n,m)=(2000,1000)$				
Estimator	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*	*Mean*	*SD*	*RMSE*	$\overline{SE}$	*CR*
					Model A ($\rho_1 = 0.1$)					
OLS*	0.9986	0.0557	0.0557	0.0560	95%	1.0019	0.0397	0.0398	0.0397	95%
OLS-S	2.9802	0.0450	1.9807	–	–	2.9806	0.0315	1.9809	–	–
IV1-S	0.7511	1.2663	1.2905	1.1517	92%	0.9048	0.6176	0.6248	0.6153	94%
IV2-S	0.7098	1.6196	1.6453	1.3411	92%	0.8952	0.6630	0.6712	0.6440	93%
2SLS-S	1.0225	0.9272	0.9274	0.9017	88%	1.0138	0.5593	0.5595	0.5729	92%
GMM-S	1.0227	0.9271	0.9274	0.9010	88%	1.0135	0.5595	0.5597	0.5727	92%
MSOLS	1.1266	0.0622	0.1410	–	–	1.0938	0.0434	0.1034	–	–
MSII	0.9061	0.0803	0.1236	–	–	0.9468	0.0505	0.0733	–	–
MSII-FM	0.8206	0.0825	0.1975	0.1065	67%	0.9053	0.0511	0.1076	0.0629	72%
PARA	0.9990	0.0609	0.0609	0.0611	95%	1.0009	0.0426	0.0426	0.0432	97%
PILS-E	1.0289	0.0718	0.0774	0.0742	94%	1.0025	0.0462	0.0463	0.0508	97%
PILS-B	1.0014	0.0635	0.0635	0.0764	98%	0.9963	0.0438	0.0439	0.0522	98%
					Model A ($\rho_1 = 0.4$)					
OLS*	0.9999	0.0293	0.0293	0.0292	94%	1.0012	0.0204	0.0205	0.0207	96%
OLS-S	1.4204	0.0451	0.4228	–	–	1.4204	0.0312	0.4215	–	–
IV1-S	1.0002	0.1163	0.1163	0.1163	96%	0.9973	0.0800	0.0801	0.0820	95%
IV2-S	0.9998	0.1197	0.1197	0.1200	95%	0.9968	0.0832	0.0833	0.0848	96%
2SLS-S	1.0028	0.1164	0.1164	0.1160	96%	0.9985	0.0799	0.0799	0.0819	96%
GMM-S	1.0028	0.1167	0.1167	0.1159	96%	0.9984	0.0798	0.0798	0.0819	96%
MSOLS	1.0166	0.0312	0.0353	–	–	1.0151	0.0219	0.0266	–	–
MSII	0.9953	0.0316	0.0320	–	–	0.9987	0.0221	0.0221	–	–
MSII-FM	0.9861	0.0317	0.0346	0.0308	92%	0.9938	0.0220	0.0229	0.0216	94%
PARA	0.9998	0.0309	0.0309	0.0307	95%	1.0010	0.0214	0.0214	0.0217	95%
PILS-E	1.0095	0.0353	0.0366	0.0324	91%	1.0035	0.0237	0.0239	0.0223	93%
PILS-B	1.0014	0.0317	0.0317	0.0310	95%	1.0012	0.0219	0.0219	0.0218	95%
					Model B ($\rho_1 = 0.1$)					
OLS*	0.9986	0.0601	0.0602	0.0613	95%	0.9996	0.0447	0.0447	0.0433	95%
OLS-S	1.4335	0.0288	0.4345	–	–	1.4343	0.0199	0.4348	–	–
IV1-S	0.9378	0.4195	0.4241	0.4168	96%	0.9830	0.2423	0.2429	0.2427	97%
IV2-S	0.9217	0.5396	0.5452	0.4821	96%	0.9805	0.2584	0.2591	0.2531	97%
2SLS-S	0.9991	0.3648	0.3648	0.3607	94%	1.0076	0.2293	0.2295	0.2326	96%
GMM-S	0.9987	0.3651	0.3651	0.3603	94%	1.0073	0.2294	0.2295	0.2325	96%
MSOLS	1.0910	0.0599	0.1090	–	–	1.0664	0.0440	0.0797	–	–
MSII	0.8618	0.1102	0.1767	–	–	0.9224	0.0617	0.0991	–	–
MSII-FM	0.8211	0.1186	0.2147	0.1308	81%	0.9081	0.0631	0.1115	0.0676	76%
PARA	1.4620	2.8409	2.8782	–	–	1.5242	4.1598	4.1927	–	–
PILS-E	1.0089	0.0702	0.0708	0.0718	95%	0.9928	0.0487	0.0493	0.0500	96%
PILS-B	0.9610	0.0728	0.0826	0.0812	96%	0.9637	0.0507	0.0624	0.0539	92%
					Model B ($\rho_1 = 0.4$)					
OLS*	0.9991	0.0369	0.0369	0.0373	96%	1.0005	0.0272	0.0271	0.0263	94%
OLS-S	1.1077	0.0282	0.1114	–	–	1.1084	0.0190	0.1101	–	–
IV1-S	0.9983	0.0683	0.0684	0.0702	96%	1.0005	0.0495	0.0495	0.0496	95%
IV2-S	0.9974	0.0706	0.0707	0.0725	95%	1.0003	0.0511	0.0511	0.0512	96%
2SLS-S	0.9991	0.0686	0.0686	0.0701	96%	1.0009	0.0494	0.0494	0.0495	95%
GMM-S	0.9991	0.0687	0.0687	0.0700	95%	1.0009	0.0494	0.0494	0.0495	95%
MSOLS	1.0405	0.0369	0.0548	–	–	1.0378	0.0266	0.0462	–	–
MSII	0.9856	0.0536	0.0555	–	–	0.9917	0.0362	0.0372	–	–
MSII-FM	0.9811	0.0551	0.0583	0.0533	93%	0.9897	0.0367	0.0381	0.0355	94%
PARA	1.1336	0.9079	0.9177	–	–	1.1168	2.0188	2.0222	–	–
PILS-E	1.0194	0.0413	0.0456	0.0453	93%	1.0111	0.0298	0.0318	0.0325	95%
PILS-B	0.9994	0.0417	0.0417	0.0554	98%	0.9992	0.0307	0.0307	0.0367	97%

3.1 Additional Results

In addition to the results reported in Section 4 of the paper, the following IV-based estimation methods are applied to the short regression with *abil* omitted:

1. Just-identification cases:

   (a) The IV for *educ* is $fatheduc$ [IV1-S]; and

   (b) The IV for *educ* is $fatheduc^2$ [IV2-S].

2. Over-identification cases where IVs for *educ* are $(fatheduc, fatheduc^2)$:

   (a) Two-stage least squares [2SLS-S]; and

   (b) Two-step generalized method of moments with 2SLS-S used as the initial estimate [GMM-S].

Table 8 presents the extended estimation results. Both $fatheduc$ and $fatheduc^2$ appear to be valid and strong IVs, because the value of the J-statistic from GMM-S is 1.5154 ($df = 1$) and $\widehat{corr}\,(educ, fatheduc) = 0.4336$, and $\widehat{corr}\,(educ, fatheduc^2) = 0.4204$. Totally different coefficient estimates and large SEs of PARA may be again due to poor linear approximation in the first step.

TABLE 8 HERE

3.2 Discussion of Inconsistencies of the Estimators

Here we discuss the estimators we consider in the empirical section in light of the different assumptions applicable to the earnings equation and to the PSID and NLS

Table 8: Estimation Results

	(1) OLS*	(2) OLS-S	(3) IV1-S	(4) IV2-S	(5) 2SLS-S	(6) GMM-S	(7) MSOLS	(8) MSII-FM	(9) PARA	(10) PILS
Education	0.0635	0.0718	0.0953	0.0908	0.0944	0.0954	0.0727	0.0685	−0.2850	0.0568
	(0.0056)	(0.0053)	(0.0159)	(0.0163)	(0.0159)	(0.0159)	(0.0059)	(0.0080)	(0.2281)	(0.0087)
Experience	0.0809	0.0818	0.0830	0.0828	0.0830	0.0826	0.0826	0.0766	0.1170	0.0818
	(0.0043)	(0.0043)	(0.0045)	(0.0044)	(0.0045)	(0.0044)	(0.0049)	(0.0065)	(0.0095)	(0.0043)
Experience^2	−0.0017	−0.0017	−0.0017	−0.0017	−0.0017	−0.0017	−0.0017	−0.0016	−0.0017	−0.0017
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Ability	0.0313	–	–	–	–	–	−0.0012	0.0029	0.2444	0.0094
	(0.0078)	(–)	(–)	(–)	(–)	(–)	(0.0027)	(0.0081)	(0.1514)	(0.0045)
Married	0.3717	0.3793	0.3844	0.3835	0.3843	0.3907	0.3799	0.3777	0.2226	0.3958
	(0.0539)	(0.0536)	(0.0536)	(0.0537)	(0.0536)	(0.0534)	(0.0535)	(0.0535)	(0.1319)	(0.0543)
Black	−0.1302	−0.1741	−0.1249	−0.1342	−0.1267	−0.1220	−0.1849	−0.1504	−0.1819	−0.1776
	(0.0323)	(0.0316)	(0.0426)	(0.0436)	(0.0427)	(0.0425)	(0.0393)	(0.0806)	(0.1146)	(0.0316)
South	−0.0921	−0.0983	−0.0814	−0.0846	−0.0821	−0.0829	−0.0979	−0.0989	−0.1372	−0.1042
	(0.0288)	(0.0287)	(0.0314)	(0.0314)	(0.0314)	(0.0314)	(0.0286)	(0.0289)	(0.1135)	(0.0287)
Urban	0.1363	0.1499	0.1278	0.1320	0.1286	0.1272	0.1538	0.1404	0.0138	0.1557
	(0.0282)	(0.0284)	(0.0332)	(0.0334)	(0.0332)	(0.0332)	(0.0297)	(0.0390)	(0.1273)	(0.0282)
Data Combination?	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Sample Size: n	2430	2430	2430	2430	2430	2430	2430	2430	2430	2430
m	–	–	–	–	–	–	1102	1102	1102	1102

Notes: The dependent variable is the log of total annual labor earnings. *Education*, *married*, *black*, *south*, and *urban* are used as the included common variables, and *age* and *south while growing up* are used as the excluded common variables. The (demeaned) *IQ score* and *KWW score* variables are used as ability measures in PSID and NLS samples, respectively. The value of the *J*-statistic from GMM-S is 1.5154 (df = 1).

samples. To save space, we focus on a reduced version of model

$$\log(earnings) = \beta_0 + \beta_1 education + \beta_2 ability + \beta_3 experience + \beta_4 experience^2 + \beta_5 married + \beta_6 black + \beta_7 south + \beta_8 urban + u,$$

where we ignore all the covariates except for the ones of interest.

As a starting point, consider the following model:

$$\log(earnings) = \beta_0 + \beta_1 education + \beta_2 ability + u. \tag{15}$$

The probability limit of the OLS estimator of β_1 from equation (15) where *ability* is unobserved is

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1^{OLS} = \beta_1 + \beta_2 \frac{cov(ability, education)}{var(education)}. \tag{16}$$

If we assume that $cov(ability, education) > 0$ and $\beta_2 > 0$, $\hat{\beta}_1^{OLS}$ for equation (15) with unobserved ability is upward-inconsistent.

Second, if there is no ability measure available in our sample, we can employ IV estimation of the wage equation (15) and it will produce a consistent estimator of β_1 as long as the instrument used – father's education – is a valid instrument for *education*.

However, if our instrument has a direct effect on the individual's wage, the IV estimator of equation (15) with unobservable ability will be inconsistent. Indeed, if the individual's wage is determined by:

$$\log(earnings) = \gamma_0 + \gamma_1 education + \gamma_2 father's\ education + u,$$

then the relation between β_1 (from equation (15) without the ability variable) and γ_1 and γ_2 can be obtained following the omitted variable formula:

$$\beta_1 = \frac{cov(father's\ education, \log(earnings))}{cov(father's\ education, education)} = \gamma_1 + \gamma_2 \frac{\lambda\sigma_f}{\sigma_e},$$

where $\lambda = corr(father's\ education, education)$, $\sigma_f^2 = var(father's\ education)$, and $\sigma_e^2 = var(education)$. In this case, the probability limit of the IV estimator of β_1 from equation (15) where *ability* is excluded from the equation is

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1^{IV} = \frac{cov(father's\ education, \log(earnings))}{cov(father's\ education, education)} = \gamma_1 + \gamma_2 \frac{\sigma_f}{\lambda\sigma_e} = \beta_1 + \gamma_2 \frac{(1-\lambda^2)\sigma_f}{\lambda\sigma_e}. \quad (17)$$

If we assume that $0 < \lambda < 1$ and $\gamma_2 > 0$, $\hat{\beta}_1^{IV}$ for equation (15) with unobserved ability is upward-inconsistent.

Third, if there is measurement error in ability under the classical errors-in-variables assumption that the correlation between the true unobserved variable and its measurement error is zero, the OLS estimator of β_1 from equation (15) with an error-ridden measure of ability will be inconsistent. Let

$$ability = ability^* + \epsilon,$$

where $ability^*$ is the unobserved error-free ability, *ability* is the observed mismeasured ability, and ϵ is the measurement error. We can show that

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1^{OLS} = \beta_1 + \beta_2 \frac{\sigma_\epsilon^2 \sigma_{ea} - \sigma_a^2 \sigma_{e\epsilon}}{\sigma_e^2 \sigma_a^2 - \sigma_{ea}^2}, \tag{18}$$

where $\sigma_\epsilon^2 = var(\epsilon)$, $\sigma_e^2 = var(education)$, $\sigma_a^2 = var(ability)$, $\sigma_{ea} = cov(education, ability)$, and $\sigma_{e\epsilon} = cov(education, \epsilon)$. If we assume that $\sigma_{ea} > 0$ and the correlation between the individual's educational level and the measurement error in his ability is zero (i.e., $\sigma_{e\epsilon} = 0$), then

$$\operatorname*{plim}_{n\to\infty} \hat{\beta}_1^{OLS} = \beta_1 + \beta_2 \frac{\sigma_\epsilon^2 \sigma_{ea}}{\sigma_e^2 \sigma_a^2 (1 - \rho_{ea}^2)}, \tag{19}$$

where $\rho_{ea} = corr(education, ability)$. Thus, $\hat{\beta}_1^{OLS}$ from equation (15) with mismeasured ability is upward-inconsistent.

Finally, note that Hirukawa and Prokhorov (2018) provide a measurement error interpretation for the MSOLS estimator of the earnings equation when *ability* is unobserved in (15). Thus, we expect the MSOLS estimator of β_1 from (15) with missing *ability* to be also upward-inconsistent following our argument above for the case of mismeasured ability.

Based on the above calculations we can make the following conclusions. First, if we use a valid instrument to estimate β_1 in (15) where *ability* is missing, then this IV estimator of β_1 is consistent while the OLS estimators of β_1 with mismeasured or omitted ability (including the MSOLS estimator) are upward-inconsistent. Second, if we think there is no measurement error in the ability measure which is available to us, then the OLS* estimator of β_1 is consistent, while the OLS-S and MSOLS are still not. Third, when the instrument for education is valid and ability is missing from equation (15) then it is, in general, unclear which one of the three (four if we count MSOLS) upward inconsistencies considered above will be larger and which smaller.

3.3 Sensitivity Analysis in PILS Estimation

A concern on PILS estimation is how sensitive it is to the choice of smoothing parameter values in nonparametric estimation of $g_2(\cdot)$. The benchmark PILS estimates in Section 4 are based on $\hat{b} = \hat{\sigma}_U (\log m/m)^{0.6}$ and $\hat{\lambda} = (\log m/m)^{0.6}$ for the beta and discrete kernels, respectively, where $\hat{\sigma}_U$ is the sample standard deviation of a converted continuous common variable U in $\mathcal{S}_2$. Numerical values of the smoothing parameters are $\hat{b}_1 = 0.0090$ for *education* (included continuous common variable), $\hat{b}_2 = 0.0022$ for *age* (excluded continuous common variable), and $\hat{\lambda} = 0.0481$ for all binary common variables.

We multiply each smoothing parameter value by $\{1/2, 2/3, 1, 3/2, 2\}$ and reestimate the regression model. Table 9 indicates how the PILS estimates change from the benchmark case ("$\left(\hat{b}_1, \hat{b}_2, \hat{\lambda}\right)$ multiplied by 1"). It can be found that even after the original smoothing parameter values are doubled or cut by half, the estimation results are qualitatively similar.

TABLE 9 HERE

Table 9: Sensitivity of PILS to Smoothing Parameter Values

$\left(\hat{b}_1, \hat{b}_2, \hat{\lambda}\right)$ *multiplied by*	1/2	2/3	1	3/2	2
Education	0.0419	0.0529	0.0568	0.0570	0.0582
	(0.0086)	(0.0070)	(0.0087)	(0.0103)	(0.0186)
Experience	0.0943	0.0732	0.0818	0.0816	0.0814
	(0.0095)	(0.0044)	(0.0043)	(0.0043)	(0.0044)
$Experience^2$	−0.0022	−0.0014	−0.0017	−0.0017	−0.0017
	(0.0004)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Ability	0.0096	0.0088	0.0094	0.0094	0.0088
	(0.0036)	(0.0033)	(0.0045)	(0.0057)	(0.0105)
Married	0.3971	0.3815	0.3958	0.3938	0.3912
	(0.0618)	(0.0510)	(0.0543)	(0.0545)	(0.0564)
Black	−0.1864	−0.2120	−0.1776	−0.1768	−0.1757
	(0.0353)	(0.0287)	(0.0316)	(0.0316)	(0.0317)
South	−0.1078	−0.1122	−0.1042	−0.1027	−0.1013
	(0.0320)	(0.0262)	(0.0287)	(0.0286)	(0.0287)
Urban	0.1310	0.1317	0.1557	0.1556	0.1550
	(0.0335)	(0.0269)	(0.0282)	(0.0282)	(0.0303)

References

[1] Dieterle, S. G., and A. Snell (2016): "A Simple Diagnostic to Investigate Instrument Validity and Heterogeneous Effects When Using a Single Instrument," *Labour Economics*, 42, 76 - 86.

[2] Eicker, F. (1963): "Asymptotic Normality and Consistency of the Least Squares Estimators for Families of Linear Regressions," *Annals of Mathematical Statistics*, 34, 447 - 456.

[3] Fang, H., M. P. Keane, and D. Silverman (2008): "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market," *Journal of Political Economy*, 116, 303 - 350.

[4] Hirukawa, M., and A. Prokhorov (2018): "Consistent Estimation of Linear Regression Models Using Matched Data," *Journal of Econometrics*, 203, 344 - 358.

[5] Horowitz, J. L., and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69, 599 - 631.

[6] White, H. (1980): "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817 - 838.