Supplement to "Heterogeneous treatment effects with mismeasured endogenous treatment"

(*Quantitative Economics*, Vol. 9, No. 3, November 2018, 1335–1370)

Takuya Ura
Department of Economics, University of California, Davis

Appendix C: Additional assumptions on the measurement error

This section considers widely-used assumptions on the measurement structure: (i) there is a positive correlation between the measurement and the truth, and (ii) the measurement error is independent of the other error in the simultaneous equation system in (1)–(3).

Assumption 22. (i) $P(T^* \neq T \mid T^*) < 1/2$. (ii) T_{t^*} *is independent of* (Y_{t^*}, T_0^*, T_1^*) *for each* $t^* = 0, 1$.

Assumption 22(i) is that the measured treatment T is equal to the true variable T^* with probability more than 50%. Assumption 22(ii) has been used in the literature on measurement error, e.g., Mahajan (2006), Lewbel (2007), and Hu (2008).

Unlike Assumption 1 itself, the combination of Assumptions 1 and 22 yields restrictions on the distribution for the observed variables.

Lemma 23. *Suppose that Assumptions* 1 *and* 22 *hold. Then*

$$\Delta f_{(Y,T)|Z}(y, 1) - \Delta f_{(Y,T)|Z}(y, 0) \geq 0.$$

Proof. As in Kitagawa (2015, Proposition 1.1) and Mourifié and Wan (2016, Theorem 1), Assumptions 1 implies the following inequalities

$$f_{(Y,T^*)|Z=0}(y, 0) \geq f_{(Y,T^*)|Z=1}(y, 0), \tag{17}$$

$$f_{(Y,T^*)|Z=0}(y, 1) \leq f_{(Y,T^*)|Z=1}(y, 1). \tag{18}$$

Assumption 22(i) implies

$$\begin{aligned} f_{(Y,T)|Z}(y, t) &= \sum_{t^*=0,1} f_{T|Y=y,T^*=t^*,Z}(t) f_{(Y,T^*)|Z}\big(y, t^*\big) \\ &= \sum_{t^*=0,1} f_{T_{t^*}|Y_{t^*}=y,T^*=t^*,Z}(t) f_{(Y,T^*)|Z}\big(y, t^*\big) \end{aligned}$$

Takuya Ura: takura@ucdavis.edu

Available at http://qeconomics.org. https://doi.org/10.3982/QE886

$$
\begin{aligned}
&= \sum_{t^*=0,1} f_{T_{t^*}}(t) f_{(Y,T^*)|Z}\big(y, t^*\big), \\
\Delta f_{(Y,T)|Z}(y,1) - \Delta f_{(Y,T)|Z}(y,0) &= \sum_{t^*=0,1} \big(f_{T_{t^*}}(1) - f_{T_{t^*}}(0)\big)\Delta f_{(Y,T^*)|Z}\big(y, t^*\big) \\
&= \big(f_{T_0}(1) - f_{T_0}(0)\big)\Delta f_{(Y,T^*)|Z}(y,0) \\
&\quad + \big(f_{T_1}(1) - f_{T_1}(0)\big)\Delta f_{(Y,T^*)|Z}(y,1) \\
&= \big(2f_{T_0}(1) - 1\big)\Delta f_{(Y,T^*)|Z}(y,0) \\
&\quad + \big(1 - 2f_{T_1}(0)\big)\Delta f_{(Y,T^*)|Z}(y,1) \\
&\geq 0,
\end{aligned}
$$

where the last inequality comes from Eq. (17)–(18) and $P(T^* \neq T \mid T^*) < 1/2$. □

The sharp identified set is characterized as follows.

THEOREM 24. *Suppose that Assumptions 1 and 22 hold, and consider an arbitrary data distribution P of (Y, T, Z). Then $\Theta_I(P) = \Theta$ if $TV_Y = 0$; otherwise*

$$
\Theta_I(P) = \left\{(1 - p_0 - p_1)\frac{\Delta E_P[Y \mid Z]}{\Delta E_P[T \mid Z]} : (p_0, p_1) \in \Omega\right\},
$$

where Ω is the set of (p_0, p_1) such that

$$
\begin{aligned}
0 \leq p_1 &\leq f_{T|Y=y,Z=1}(0), \\
0 \leq p_0 &\leq f_{T|Y=y,Z=0}(1), \\
p_1 \Delta f_{Y|Z}(y) &\geq \Delta f_{(Y,T)|Z}(y,0), \\
p_0 \Delta f_{Y|Z}(y) &\leq \Delta f_{(Y,T)|Z}(y,1)
\end{aligned}
$$

for every y.

PROOF. Define

$$
\begin{pmatrix} a_{00}(y) & a_{01}(y) \\ a_{10}(y) & a_{11}(y) \end{pmatrix} = \begin{pmatrix} f_{(Y,T)|Z=0}(y,0) & f_{(Y,T)|Z=1}(y,0) \\ f_{(Y,T)|Z=0}(y,1) & f_{(Y,T)|Z=1}(y,1) \end{pmatrix},
$$

$$
\begin{pmatrix} b_{00}(y) & b_{01}(y) \\ b_{10}(y) & b_{11}(y) \end{pmatrix} = \begin{pmatrix} f_{(Y,T^*)|Z=0}(y,0) & f_{(Y,T^*)|Z=1}(y,0) \\ f_{(Y,T^*)|Z=0}(y,1) & f_{(Y,T^*)|Z=1}(y,1) \end{pmatrix},
$$

and define $p_0 = f_{T|T^*=0}(1)$ and $p_1 = f_{T|T^*=1}(0)$. Since $f_{(Y,T)|Z}(y,t) = \sum_{t^*=0,1} f_{T_{t^*}}(t) \times f_{(Y,T^*)|Z}(y,t^*)$ from Assumption 22(i), it follows that

$$
\begin{pmatrix} a_{00}(y) & a_{01}(y) \\ a_{10}(y) & a_{11}(y) \end{pmatrix} = \begin{pmatrix} 1 - p_0 & p_1 \\ p_0 & 1 - p_1 \end{pmatrix} \begin{pmatrix} b_{00}(y) & b_{01}(y) \\ b_{10}(y) & b_{11}(y) \end{pmatrix}.
$$

Assumption 22(i) implies that the matrix $\left(\begin{smallmatrix} 1-p_0 & p_1 \\ p_0 & 1-p_1 \end{smallmatrix}\right)$ is invertible. Thus

$$\begin{pmatrix} b_{00}(y) & b_{01}(y) \\ b_{10}(y) & b_{11}(y) \end{pmatrix} = \frac{1}{1-p_0-p_1} \times \begin{pmatrix} (1-p_1)a_{00}(y) - p_1 a_{10}(y) & (1-p_1)a_{01}(y) - p_1 a_{11}(y) \\ -p_0 a_{00}(y) + (1-p_0)a_{10}(y) & -p_0 a_{01}(y) + (1-p_0)a_{11}(y) \end{pmatrix}$$

and

$$\theta(P) = (1-p_0-p_1)\frac{\Delta E[Y \mid Z]}{\Delta E[T \mid Z]},$$

because

$$\begin{aligned} \Delta E\big[T^* \mid Z\big] &= f_{T^*|Z=1}(1) - f_{T^*|Z=0}(1) \\ &= \int b_{11}(y)\,dy - \int b_{10}(y)\,dy \\ &= \frac{p_0 \int \big(a_{00}(y) - a_{01}(y)\big)\,dy + (1-p_0)\int \big(a_{11}(y) - a_{10}(y)\big)\,dy}{1-p_0-p_1} \\ &= \frac{-p_0 \Delta E[1-T \mid Z] + (1-p_0)\Delta E[T \mid Z]}{1-p_0-p_1} \\ &= \frac{\Delta E[T \mid Z]}{1-p_0-p_1}. \end{aligned}$$

In the rest of the proof, I am going to show that the sharp identified set for (p_0, p_1) is Ω.

First, I am going to show that the identified set for (p_0, p_1) is a subset of Ω. As in Kitagawa (2015, Proposition 1.1) and Mourifié and Wan (2016, Theorem 1), Assumption 1 implies the following inequalities

$$\begin{aligned} f_{(Y,T^*)|Z=0}(y,0) \geq f_{(Y,T^*)|Z=1}(y,0) \geq 0, \\ 0 \leq f_{(Y,T^*)|Z=0}(y,1) \leq f_{(Y,T^*)|Z=1}(y,1). \end{aligned}$$

In the notation of this proof,

$$\begin{aligned} (1-p_1)a_{00}(y) - p_1 a_{10}(y) \geq (1-p_1)a_{01}(y) - p_1 a_{11}(y) \geq 0, \\ 0 \leq -p_0 a_{00}(y) + (1-p_0)a_{10}(y) \leq -p_0 a_{01}(y) + (1-p_0)a_{11}(y). \end{aligned}$$

By some algebraic operations,

$$\begin{aligned} p_1 f_{Y|Z=1}(y) &\leq f_{(Y,T)|Z=1}(y,0), \\ p_0 f_{Y|Z=0}(y) &\leq f_{(Y,T)|Z=0}(y,1), \\ p_1 \Delta f_{Y|Z}(y) &\geq \Delta f_{(Y,T)|Z}(y,0), \\ p_0 \Delta f_{Y|Z}(y) &\leq \Delta f_{(Y,T)|Z}(y,1). \end{aligned}$$

Then, I am going to show that Ω is included in the identified set for (p_0, p_1). Let $(\tilde{p}_0, \tilde{p}_1)$ be any element of Ω. Then define

$$\begin{pmatrix} \tilde{f}_{(Y,T^*)|Z=0}(y,0) & \tilde{f}_{(Y,T^*)|Z=1}(y,0) \\ \tilde{f}_{(Y,T^*)|Z=0}(y,1) & \tilde{f}_{(Y,T^*)|Z=1}(y,1) \end{pmatrix} = \begin{pmatrix} 1-\tilde{p}_0 & \tilde{p}_1 \\ \tilde{p}_0 & 1-\tilde{p}_1 \end{pmatrix}^{-1} \begin{pmatrix} a_{00}(y) & a_{01}(y) \\ a_{10}(y) & a_{11}(y) \end{pmatrix}.$$

Some calculations yield

$$\tilde{f}_{(Y,T^*)|Z=0}(y,0) \geq \tilde{f}_{(Y,T^*)|Z=1}(y,0) \geq 0,$$
$$0 \leq \tilde{f}_{(Y,T^*)|Z=0}(y,0) \leq \tilde{f}_{(Y,T^*)|Z=1}(y,0).$$

This is a sufficient condition for $\tilde{f}_{(Y,T^*)|Z}$ to be consistent with Assumptions 1, which is shown in Kitagawa (2015, Proposition 1.1) and Mourifié and Wan (2016, Theorem 1). Thus $(\tilde{p}_0, \tilde{p}_1)$ belongs to the identified set for (p_0, p_1). □

Appendix D: Compliers-defiers-for-marginals condition

This section demonstrates that a variant of Theorem 4 still holds under a weaker condition than the deterministic monotonicity condition in Assumption 1(ii). I consider the following assumption.

Assumption 25. (i) *For each* $t^* = 0, 1$, Z *is independent of* $(T_{t^*}, Y_{t^*}, T_0^*, T_1^*)$. (ii) *There is a subset* C_F *of* $\{T_1^* > T_0^*\}$ *such that*

$$P(C_F) = P\big(T_1^* < T_0^*,\big),$$
$$f_{(Y_0,T_0)|C_F} = f_{(Y_0,T_0)|T_1^*<T_0^*},$$
$$f_{(Y_1,T_1)|C_F} = f_{(Y_1,T_1)|T_1^*<T_0^*}.$$

(iii) $0 < P(Z=1) < 1$.

Assumption 25(ii) imposes the compliers-defiers-for-marginals condition (de Chaisemartin (2016)) on the joint distribution of (Y_{t^*}, T_{t^*}). Under this assumption, Theorem 2.1 of de Chaisemartin (2016) shows that

$$E[Y_0 - Y_1 \mid C_V] = \frac{\Delta E[Y \mid Z]}{\Delta E\big[T^* \mid Z\big]},$$

where $C_V = \{T_1^* > T_0^*\} \setminus C_F$.[12]

[12]In fact he uses a weaker condition than the compliers-defiers-for-marginals condition to establish this equality. I use the compliers-defiers-for-marginals condition here, because it makes the characterization in 26 exactly the same to Theorem 2.1.

THEOREM 26. *Suppose that Assumption* 1 *holds, and consider an arbitrary data distribution P of (Y, T, Z). The identified set $\tilde{\Theta}_I(P)$ for $E[Y_0 - Y_1 \mid C_V]$ is characterized in the same way as Theorem* 4: *$\tilde{\Theta}_I(P) = \Theta$ if $TV_{(Y,T)} = 0$; otherwise,*

$$\tilde{\Theta}_I(P) = \begin{cases} \left[\Delta E_P[Y \mid Z], \dfrac{\Delta E_P[Y \mid Z]}{TV_{(Y,T)}}\right] & \text{if } \Delta E_P[Y \mid Z] > 0, \\ \{0\} & \text{if } \Delta E_P[Y \mid Z] = 0, \\ \left[\dfrac{\Delta E_P[Y \mid Z]}{TV_{(Y,T)}}, \Delta E_P[Y \mid Z]\right] & \text{if } \Delta E_P[Y \mid Z] < 0. \end{cases}$$

PROOF. Since Theorem 4 gives the sharp identified set under a stronger assumption of this theorem, the identified set $\tilde{\Theta}_I(P)$ in this theorem should be equal to or larger than the set in Theorem 4. As a result, it suffices to show that $\tilde{\Theta}_I(P)$ is a subset of the set in Theorem 4. By the Assumption 25(ii),

$$\begin{aligned} f_{(Y,T)|Z=0}(y,t) &= P\big(T_1^* = T_0^* = 1 \mid Z=0\big) f_{(Y,T)|T_1^*=T_0^*=1, Z=0}(y,t) \\ &\quad + P\big(T_1^* = T_0^* = 0 \mid Z=0\big) f_{(Y,T)|T_1^*=T_0^*=0, Z=0}(y,t) \\ &\quad + P(C_F \mid Z=0) f_{(Y,T)|C_F, Z=0}(y,t) \\ &\quad + P(C_V \mid Z=0) f_{(Y,T)|C_V, Z=0}(y,t) \\ &\quad + P\big(T_1^* < T_0^* \mid Z=0\big) f_{(Y,T)|T_1^*<T_0^*, Z=0}(y,t) \\ &= P\big(T_1^* = T_0^* = 1\big) f_{(Y_1,T_1)|T_1^*=T_0^*=1}(y,t) \\ &\quad + P\big(T_1^* = T_0^* = 0\big) f_{(Y_0,T_0)|T_1^*=T_0^*=0}(y,t) \\ &\quad + P(C_F) f_{(Y_0,T_0)|C_F}(y,t) \\ &\quad + P(C_V) f_{(Y_0,T_0)|C_V}(y,t) \\ &\quad + P\big(T_1^* < T_0^*\big) f_{(Y_1,T_1)|T_1^*<T_0^*}(y,t), \\ f_{(Y,T)|Z=1}(y,t) &= P\big(T_1^* = T_0^* = 1 \mid Z=1\big) f_{(Y,T)|T_1^*=T_0^*=1, Z=1}(y,t) \\ &\quad + P\big(T_1^* = T_0^* = 0 \mid Z=1\big) f_{(Y,T)|T_1^*=T_0^*=0, Z=1}(y,t) \\ &\quad + P(C_F \mid Z=1) f_{(Y,T)|C_F, Z=1}(y,t) \\ &\quad + P(C_V \mid Z=1) f_{(Y,T)|C_V, Z=1}(y,t) \\ &\quad + P\big(T_1^* < T_0^* \mid Z=1\big) f_{(Y,T)|T_1^*<T_0^*, Z=1}(y,t) \\ &= P\big(T_1^* = T_0^* = 1\big) f_{(Y_1,T_1)|T_1^*=T_0^*=1}(y,t) \\ &\quad + P\big(T_1^* = T_0^* = 0\big) f_{(Y_0,T_0)|T_1^*=T_0^*=0}(y,t) \\ &\quad + P(C_F) f_{(Y_1,T_1)|C_F}(y,t) \\ &\quad + P(C_V) f_{(Y_1,T_1)|C_V}(y,t) \\ &\quad + P\big(T_1^* < T_0^*\big) f_{(Y_0,T_0)|T_1^*<T_0^*}(y,t), \end{aligned}$$

$$\begin{aligned}\Delta f_{(Y,T)|Z}(y,t) &= P(C_F)\big(f_{(Y_1,T_1)|C_F}(y,t) - f_{(Y_0,T_0)|C_F}(y,t)\big)\\ &\quad + P(C_V)\big(f_{(Y_1,T_1)|C_V}(y,t) - f_{(Y_0,T_0)|C_V}(y,t)\big)\\ &\quad + P(T_1^* < T_0^*)\big(f_{(Y_0,T_0)|T_1^*<T_0^*}(y,t) - f_{(Y_1,T_1)|T_1^*<T_0^*}(y,t)\big)\\ &= P(C_V)\big(f_{(Y_1,T_1)|C_V}(y,t) - f_{(Y_0,T_0)|C_V}(y,t)\big).\end{aligned}$$

Based on the definition of the total variation distance,

$$TV_{(Y,T)} = P(C_V)\frac{1}{2}\sum_{t=0,1}\int \big|f_{(Y_1,T_1)|C_V}(y,t) - f_{(Y_0,T_0)|C_V}(y,t)\big|\, d\mu_Y(y),$$

and therefore

$$TV_{(Y,T)} \le P(C_V) \le 1.$$

Since $P(C_V) = \Delta E[T \mid Z]$,

$$TV_{(Y,T)} \le \Delta E[T \mid Z] \le 1.$$

This concludes that $E[Y_0 - Y_1 \mid C_V]$ is included in

$$\begin{cases}\left[\Delta E[Y \mid Z], \dfrac{\Delta E[Y \mid Z]}{TV_{(Y,T)}}\right] & \text{if } \Delta E[Y \mid Z] > 0,\\ \{0\} & \text{if } \Delta E[Y \mid Z] = 0,\\ \left[\dfrac{\Delta E[Y \mid Z]}{TV_{(Y,T)}}, \Delta E[Y \mid Z]\right] & \text{if } \Delta E[Y \mid Z] < 0.\end{cases}$$

□

Appendix E: Numerical example and Monte Carlo simulations

This section considers a numerical example to illustrate the theoretical properties in the previous section. I consider the following data generating process:

$$\begin{aligned} Z &\sim \text{Bernoulli}(0.5),\\ T^* &= 1\{-3/4 + 1/2Z + U_1 \ge 0\},\\ Y &= 2T^* + \Phi(U_2),\\ T &= T^* + \big(1 - 2T^*\big)1\{U_3 \le \gamma\},\end{aligned}$$

where Φ is the standard normal cdf and, conditional on Z, (U_1, U_2, U_3) is drawn from the Gaussian copula with the correlation matrix

$$\begin{pmatrix} 1 & 0.25 & 0.25\\ 0.25 & 1 & 0.25\\ 0.25 & 0.25 & 1\end{pmatrix}.$$

I set $\gamma = 0, 0.2, 0.4$, which captures the degree of the misclassification. In this design, the treatment variable is endogenous since U_1 and U_2 are correlated. In addition, the misclassification is endogenous in that U_2 and U_3 are correlated.

TABLE 5. Population parameters (Theorem 4).

γ	LATE	Identified Set	Wald Estimand
0	2.00	$[1, 2.00]$	2.00
0.2	2.00	$[1, 2.41]$	3.01
0.4	2.00	$[1, 2.64]$	8.72

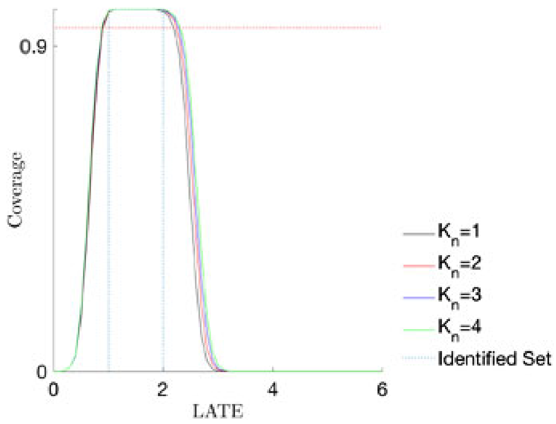

FIGURE 3. Coverage of the confidence interval (Theorem 4) for $\gamma = 0$.

Table 5 lists the three population objects: the local average treatment effect, the Wald estimand, and the identified set for the local average treatment effect. Note that, unless $\gamma = 0$, the distribution for (Y, T, Z) violates the conditions in (5). When there is no measurement error, the sharp upper bound is equal to the Wald estimand, which is the case for $\gamma = 0$. When there is a measurement error, the sharp upper bound for the local average treatment effect can be smaller than the Wald estimand.

In order to focus on the finite sample properties of the test $\mathbb{1}\{T(\theta, \pi) > c(\alpha, \theta, \pi)\}$, I only evaluate coverage probabilities given $\pi = 0.5$ for various value of θ. The partition of grids is equally spaced over $\mathbf{Y}$ with the number the partitions $K_n = 1, \ldots, 4$. Coverage probabilities are calculated as how often the 95% confidence interval includes a given parameter value out of 1000 simulations. The sample size is $n = 500$ for Monte Carlo simulations. I use 1000 bootstrap repetitions to construct critical values. I set $\beta = 0.1\%$ for the moment selection.

Figures 3–5 describe the coverage probabilities of the confidence intervals for each parameter value. When the degree of measurement error is zero $(\gamma = 0)$, the power for the confidence interval with $K_n = 1$ has a slightly better performance than those with $K_n \geq 2$. It can be because the number of moment inequalities are larger for $K_n \geq 2$ and then the critical value is bigger. As the degree of measurement error becomes larger, the power for the confidence intervals with $K_n \geq 2$ becomes better than that with $K_n = 1$. It is a result of the fact that the sharp upper bound for the local average treatment effect is smaller than the Wald estimand.

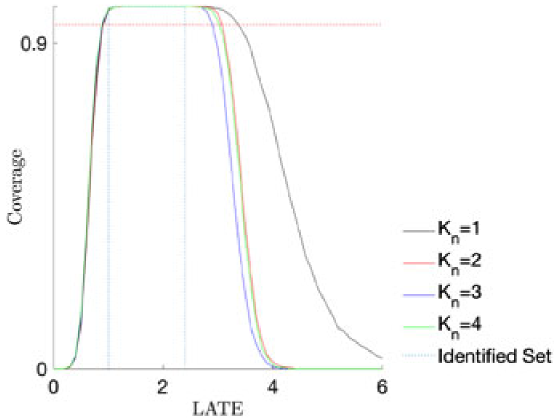

FIGURE 4. Coverage of the confidence interval (Theorem 4) for $\gamma = 0.2$.

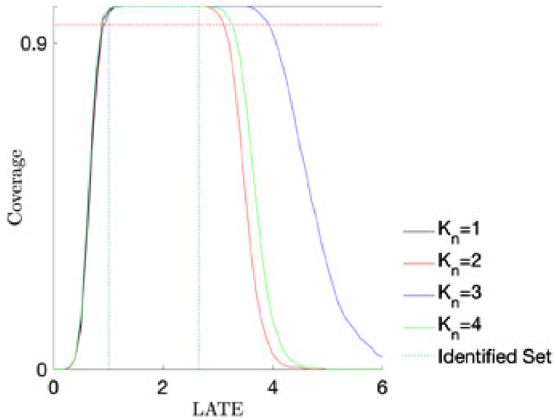

FIGURE 5. Coverage of the confidence interval (Theorem 4) for $\gamma = 0.4$.

Next, I investigate the identifying power of an additional measurement.

$$R = T^* + (1 - 2T^*)1\{U_4 \leq \gamma\},$$

where (U_1, U_2, U_3, U_4) is drawn from the Gaussian copula with the correlation matrix

$$\begin{pmatrix} 1 & 0.25 & 0.25 & 0.25 \\ 0.25 & 1 & 0.25 & 0.25 \\ 0.25 & 0.25 & 1 & 0.25 \\ 0.25 & 0.25 & 0.25 & 1 \end{pmatrix}.$$

Table 6 lists the three population objects: the local average treatment effect, the Wald estimand, and the identified set for the local average treatment effect. Figures 6–8 describe the coverage probabilities of the confidence intervals for each parameter value. The comparison among different K_n's are similar to the previous figures.

TABLE 6. Population parameters with using R (Theorem 9).

γ	LATE	Identified Set	Wald Estimand
0	2.00	[1, 2.00]	2.00
0.2	2.00	[1, 2.26]	3.01
0.4	2.00	[1, 2.62]	8.72

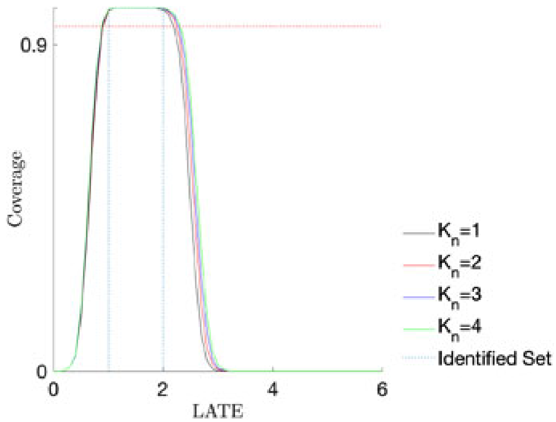

FIGURE 6. Coverage of the confidence interval with using R (Theorem 9) for $\gamma = 0$.

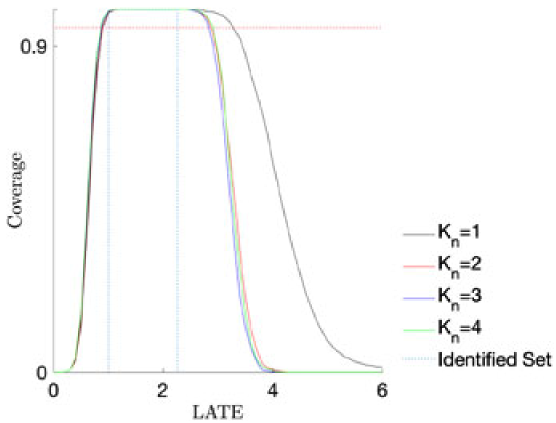

FIGURE 7. Coverage of the confidence interval with using R (Theorem 9) for $\gamma = 0.2$.

Last, I consider the dependence between measurement error and instrumental variable, as in Section 3.4. Table 7 lists the three population objects and Figures 9–11 describe the coverage probabilities of the confidence intervals. Since they do not use any information from the measured treatment T, the identified sets and the confidence intervals show that the upper bounds on the local average treatment effect is larger than those under the independence between measurement error and instrumental variable. The difference becomes smaller when the degree of the measurement error is larger. It

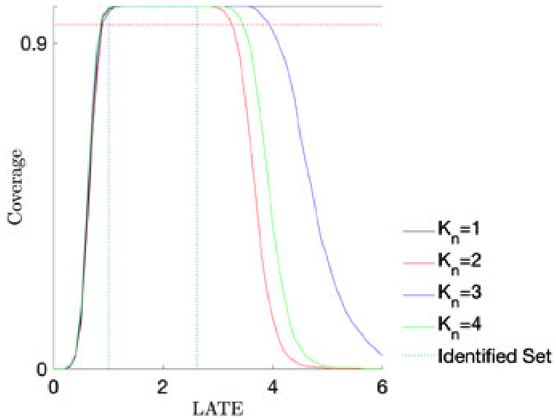

FIGURE 8. Coverage of the confidence interval with using R (Theorem 9) for $\gamma = 0.4$.

TABLE 7. Population parameters without T (Theorem 11).

γ	LATE	Identified Set	Wald Estimand
0	2.00	[1, 2.67]	2.00
0.2	2.00	[1, 2.67]	3.01
0.4	2.00	[1, 2.68]	8.72

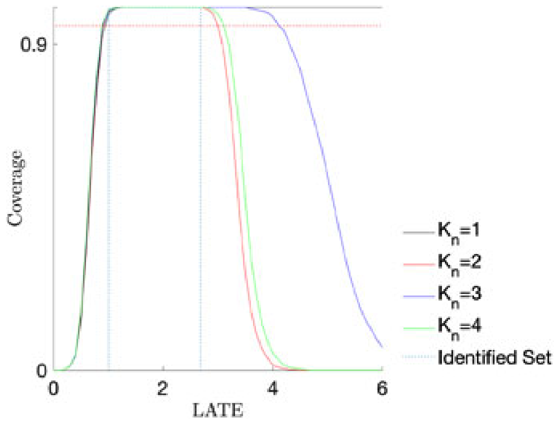

FIGURE 9. Coverage of the confidence interval without T (Theorem 11) for $\gamma = 0$.

can be considered as the result that, when the misclassification happens too often, the measured treatment T has only little information about the true treatment and therefore there is a small difference between the identified sets.

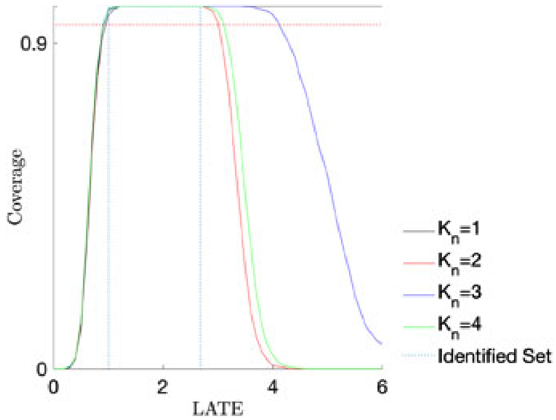

FIGURE 10. Coverage of the confidence interval without T (Theorem 11) for $\gamma = 0.2$.

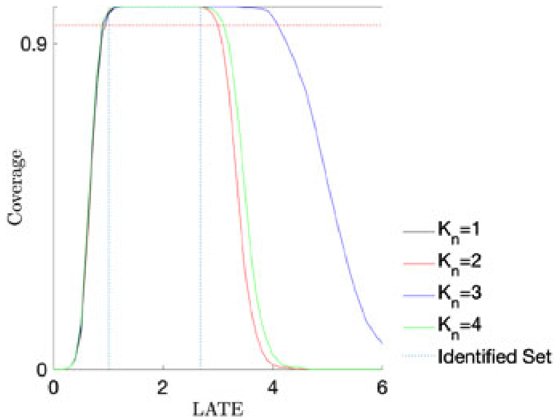

FIGURE 11. Coverage of the confidence interval without T (Theorem 11) for $\gamma = 0.4$.

REFERENCES

de Chaisemartin, C. (2016), "Tolerating defiance? Local average treatment effects without monotonicity." Quantitative Economics. (forthcoming). [4]

Hu, Y. (2008), "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution." *Journal of Econometrics*, 144, 27–61. [1]

Kitagawa, T. (2015), "A test for instrument validity." *Econometrica*, 83, 2043–2063. [1, 3, 4]

Lewbel, A. (2007), "Estimation of average treatment effects with misclassification." *Econometrica*, 75, 537–551. [1]

Mahajan, A. (2006), "Identification and estimation of regression models with misclassification." *Econometrica*, 74, 631–665. [1]

Mourifié, I. Y. and Y. Wan (2016), "Testing local average treatment effect assumptions." Review of Economics and Statistics. (forthcoming). [1, 3, 4]

Co-editor Rosa L. Matzkin handled this manuscript.

Manuscript received 19 May, 2017; final version accepted 19 February, 2018; available online 14 March, 2018.