Online Appendix for "Simple Tests for Selection: Learning More from Instrumental Variables"

Dan A. Black
University of Chicago

Joonhwi Joo
University of Texas at Dallas

Robert LaLonde
University of Chicago

Jeffrey A. Smith
University of Wisconsin at Madison

Evan J. Taylor
University of Arizona

July 28, 2022

B Online Appendix: Monte Carlo Study

B.1 Data generating process

This appendix presents a Monte Carlo (MC) analysis that illustrates the finite sample behavior of the test described in the main text of the paper. In particular, we document how the performance of the test changes as we vary (i) instrument strength, (ii) sample size, and (iii) the extent of selection on unobserved variables. The reader will recall that our test involves estimating linear outcome equations that include the instrument as a covariate, with our test statistic consisting of the estimated coefficient on the instrument. As we estimate the linear outcome equation via OLS, the test statistic has all the usual statistical properties (e.g., unbiasedness and asymptotic normality), leading us to focus our analysis not on the sampling behavior of the test statistic itself but instead on the size and power of the test.

For our MC analyses, we specify a generalized Roy model as the population data generating process. In this model, participation status D_i may depend on the instrument Z_i and on an unobserved variable, which in turn may have a non-zero correlation with the unobserved components of the outcome equations. Thus, it provides us with a rich structure that can embody varying amounts of selection on unobserved variables.

The data are generated from the following data generating process:

$$Y_{1i} = 3 + \epsilon_{1i} \tag{B1}$$

$$Y_{0i} = 1 + \epsilon_{0i} \tag{B2}$$

$$D_i^* = -\frac{\gamma^d}{3} + Z_i\gamma^d + \epsilon_{di} \tag{B3}$$

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0, \rightarrow \text{only } Y_{1i} \text{ is observed} \\ 0 & \text{if } D_i^* < 0, \rightarrow \text{only } Y_{0i} \text{ is observed} \end{cases}.$$

Equations (B1) and (B2) correspond to equations (8) and (9) in the main text. In this model, γ^d captures the strength of the instrument, with higher (absolute) values of γ^d indicating a substantively stronger instrument. We assume that $(\epsilon_{0i}, \epsilon_{1i}, \epsilon_{di}) \perp\!\!\!\perp Z_i$ and draw Z_i from a binomial distribution with $\Pr(Z_i = 0) = \Pr(Z_i = 1) = 0.5$. The error terms follow a joint normal distribution with the covariance structure:

$$\begin{pmatrix} \epsilon_{0i} \\ \epsilon_{1i} \\ \epsilon_{di} \end{pmatrix} \sim iid\,\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \qquad \mathbf{\Sigma} = \begin{pmatrix} \sigma_0^2 & -\rho\sigma_0\sigma_1 & -\rho\sigma_0\sigma_d \\ -\rho\sigma_0\sigma_1 & \sigma_1^2 & \rho\sigma_1\sigma_d \\ -\rho\sigma_0\sigma_d & \rho\sigma_1\sigma_d & \sigma_d^2 \end{pmatrix}.$$

We set $\sigma_0 = \sigma_1 = 2$. Following standard practice in probit models, we normalize $\sigma_d = 1$.

The parameter ρ, which we vary in our MC simulations, determines the correlations among the error terms, with $corr(\epsilon_{1i}, \epsilon_{di}) = \rho$ and $corr(\epsilon_{0i}, \epsilon_{di}) = corr(\epsilon_{1i}, \epsilon_{0i}) = -\rho$. Implicitly, ρ governs the degree of selection on unobserved variables. Setting $\rho = 0$ implies no selection on unobserved variables, so that $\alpha_0 = 0$ and $\alpha_1 = 0$. In contrast, a $\rho \neq 0$ implies that $\alpha_0 \neq 0$ and $\alpha_1 \neq 0$. Put differently, $\rho = 0$ implies that the null hypotheses tested by our test hold in the population, while $\rho \neq 0$ implies that those null hypotheses do not hold in the population. A larger (absolute) value of ρ implies substantively stronger selection on unobserved variables.

To see how the sign of the selection effect plays out, assume $\rho > 0$ so that $corr(\epsilon_{1i}, \epsilon_{di}) > 0$. Now observe that in equation (B3), $Z_i = 1$ implies that, at the margin, $D_i = 1$ for observations with lower values of ϵ_{di}. In notation, $E[\epsilon_{1i}|Z_i = 1] < 0$ in this case, indicating that the instrument induces negative selection into treatment on (the unobserved component of) Y_{1i} in equation (B1). A parallel argument, but with opposite signs, holds for selection on the unobserved component of Y_{0i} in equation (B2). Note that having a single ρ parameter, rather than positing separate correlations between ϵ_{di} and each of ϵ_{1i} and ϵ_{0i}, imposes important substantive restrictions on the model. In particular, it rules out patterns like the one we find in our analysis of the data from Angrist and Evans (1998) in Section 5.1, with selection into treatment based on (the unobserved component of) Y_{0i} but not (the unobserved component of) Y_{1i}. We make this restriction solely to reduce the number of parameters we have to keep track of in the MC analysis.

We do not include exogenous covariates X_i like those in equations (13) and (14) in the main text in our data generating process because we do not need them to make the points we want to make with the MC analysis given our focus on the size and power of our test. Indeed, given our assumptions on Z_i, including it would not affect the asymptotic distribution of the estimates of the auxilliary functions $g_0(X_i)$ and $g_1(X_i)$. In short, omitting covariates keeps things simple without losing any substance. More broadly, we chose not to undertake an "empirical Monte Carlo" analysis like that in Huber, Lechner and Wunsch (2013) as we want to emphasize the applicability of our test in many quite distinct empirical contexts.

B.2 Monte Carlo details

We conduct two sets of MC analyses to examine the performance of our proposed test. We set the nominal size of the test to 0.05 throughout, which is to say that we reject the null in MC samples in which the estimated test statistic exceeds 1.96 in absolute value. For each set of parameter values and sample size, we draw 1000 Monte Carlo samples from an artificial population of 100,000,000 observations created based on the DGP described above. In each Monte Carlo sample, we estimate equations (8) and (9) and perform our test as described in the main text. We also estimate the linear probability model analogue to our participation probit, i.e., a linear regression of D_i on Z_i, which provides us with first-stage F-statistics.

In the first set of MC analyes, we fix ρ and vary the sample size and the strength of the instrument as embodied in γ_d. Specifically, we fix $\rho = 0.6$, implying quite strong selection on unobserved variables. Of course, in most actual empirical exercises the researcher would reduce the extent of selection on unobserved variables by including observed variables, as we do not for the reasons noted above; see also the related discussion around equation (36) in Section 6.2 of the main text. Setting $\rho = 0.6$ implies that our nulls that $\alpha_0 = 0$ and $\alpha_1 = 0$ do not hold (i.e. are false) in the population because $\rho \neq 0$. We then vary $\gamma^d \in \{0.00, 0.05, 0.15, 0.25, 0.50, 1.50, 2.50\}$, thereby varying instrument strength from none—i.e., abject failure of the second part of the (EI) assumption in Section 3—to levels of instrument strength rarely seen outside of randomized control trials. To provide a scaling more familiar to applied researchers, we present means of realized first-stage F statistics among our MC results. Finally, we repeat our analyses for sample sizes in $N \in \{100, 500, 1{,}000, 5{,}000, 10{,}000, 50{,}000\}$, which captures the range of samples sizes used in most social science empirical work.

In the second set of MC analyses, we fix the sample size N and vary the instrument strength as embodied in γ_d and the substantive importance of selection on observed variables as embodied in ρ. We fix the sample size first at $N = 1,000$ and then at $N = 10,000$. We vary $\gamma^d \in \{0.00, 0.05, 0.15, 0.25, 0.50, 1.50, 2.50\}$, just as in the first set of simulations. We examine $\rho \in \{0.0, 0.2, 0.4, 0.6\}$. This varies the degree of selection on unobserved variables from none, which implies that the nulls $\alpha_0 = 0$ and $\alpha_1 = 0$ hold in the population, to the quite substantial amount assumed in the first set of MC analyses.

B.3 Results

Tables B1 and B2 report the results of the MC analyses. Each block of each table consists of nine rows, with the values in each row calculated using the estimates from the 1,000 independent Monte Carlo samples for that block. The first row, labeled "Mean $\hat{\alpha}_0$," gives the mean of the estimated coefficients on Z_i or, equivalently, the mean of the estimated test statistic. We include this row mainly to confirm that our simulations work as intended. The second row, labeled "Std. Dev. $\hat{\alpha}_0$," presents the standard deviation of the estimated coefficients on Z_i. It gives a more continuous sense of the variability in the test statistic across MC samples than the rejection probability provides. The third row, labeled "Mean N_0," gives the average number of untreated units in the Monte Carlo samples for each block. We picked the parameters of our DGP to generate treatment probabilities around 0.5, so this row again serves mainly to confirm the good behavior of our code.

The fourth row, labeled "Prob. Rej. $H_0 : \alpha_0 = 0$" indicates the fraction of MC samples in the block for which the null is rejected at the 0.05 level. In blocks wherein $\rho = 0.0$, so that the null that $\alpha_0 = 0$ holds in the population, this rejection probability represents the realized size of the test, i.e., the realized probability of a Type I error. For blocks wherein $\rho > 0$, so that the null that $\alpha_0 = 0$ does not hold in the population, this probability represents the realized power of the test, i.e., one minus the probability of a Type II error.

The fifth through eighth rows in each block repeat the same statistics as in the first four rows but for the $D_i = 1$ units in each MC sample. Finally, The ninth and final row in each block, labeled "Mean 1st Stage F-stat." offers exactly that, namely the mean of the F-statistics obtained from estimating linear probability models of D_i on Z_i using the MC samples for the block. The first-stage F-statistic is common to the treated and untreated units and so appears only once.

Table B1 reports the results for the first set of Monte Carlo analyses, while Table B2 reports the results from the second set. Taking the tables together reveals several general findings. First, the realized size of the test under the null of no selection on unobserved variables equals approximately 0.05. Second, the power of the test increases in the sample size, holding constant instrument strength and the amount of selection on unobserved variables. Third, the power of the test increases in instrument strength holding constant sample size and the amount of selection on unobserved variables. Fourth, and finally, the power of the test increases in the degree of selection on unobserved variables, holding constant sample size and instrument strength.

In addition to the completely unsurprising qualitative findings just described, we can also say a bit about the quantitative levels of power, keeping in mind our decision to adopt a very simple DGP that does not closely correspond to any particular empirical context. We frame this discussion in part around the mean first-stage F-statistic, keeping in mind that it increases in both the sample size (and thus across columns in Table B1 and across panels in Table B2) and in the value of γ_d, and thus with blocks in both Tables B1 and B2.

In Table B1, a mean first-stage F-statistic over 100, as demanded by Lee, McCrary, Moreira, and Porter (2021), ensures statistical power close to 1.0. A comparison with Table B2 shows that this relationship need not hold when $\rho < 0.6$, as in the case with $N = 1,000$, $\gamma_d = 1.5$ and $\rho = 0.2$. Even in this case, though, the power reaches a respectable, if not overwhelming, level of 0.47.

First-stage F-statistics exceeding 100 are thin on the ground outside of randomized control trials and nearly sharp regression discontinuity designs. What happens in cases where the mean of the first-stage F-statistics lies above the traditional rule-of-thumb level of 10.0 but below 100.0? Here the realized power depends very much on the extent of selection on unobserved variables and on just where the realized first-stage F-statistic lies within the range from 10 to 100. To see the first point, consider the case in Table B2 with $N = 1,000$ and $\gamma_d = 0.5$. Here the mean first-stage F-statistic equals around 41. With $\rho = 0.2$, this implies a realized power of only about 0.11 relative to both nulls, while when ρ rises to 0.6, the realized power rises to around 0.67. To see the second point, compare the block just considered to the one that precedes it in Table B2, wherein $\gamma_d = 0.25$ rather than $\gamma_d = 0.5$.

In that block, the mean first-stage F-statistic equals about 11 and the realized power ranges from 0.07 when $\rho = 0.2$ to 0.21 when $\rho = 0.6$.

Keeping in mind that we have not specialized our Monte Carlo analysis to any specific empirical context nor have we included any covariates in it, we view it as sending a relatively clear message. In cases with a strong first-stage and substantial amounts of selection on unobserved variables, our test does really well in terms of statistical power. In cases with a meaningful first stage but not an overpowering one, the power of the test increases in the degree of selection on unobserved variables. While generally not close to one, in many such cases it is also quite far from 0.05, suggesting that the benefits of our inexpensive-to-implement test will still exceed its costs for most researchers. Not unrelated, in these intermediate cases rejections are, in an important sense, more informative than failures to reject. Put differently, our test will yield more false negatives than false positives. Finally, in cases with a weak first stage, the researcher has bigger troubles to worry about than the low power of our test.

Table B1(a): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Different Sample Sizes (1/3)

γ^d			N =100	N =500	N =1,000	N =5,000	N =10,000	N =50,000
0	Y_0	Mean $\hat{\alpha}_0$	-0.02	0.00	0.00	0.00	0.00	0.00
		Std. Dev. $\hat{\alpha}_0$	(0.50)	(0.22)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_0	49.89	250.18	500.82	2,498.77	5,000.55	25,000.99
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.046	0.049	0.051	0.071	0.050	0.055
	Y_1	Mean $\hat{\alpha}_1$	-0.01	0.00	0.00	0.00	0.00	0.00
		Std. Dev. $\hat{\alpha}_1$	(0.49)	(0.22)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_1	50.11	249.82	499.18	2,501.23	4,999.45	24,999.01
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.044	0.053	0.051	0.052	0.048	0.058
	Mean 1st Stage F-stats		1.07	0.98	1.07	0.97	1.05	1.01
0.05	Y_0	Mean $\hat{\alpha}_0$	0.07	0.04	0.04	0.04	0.04	0.04
		Std. Dev. $\hat{\alpha}_0$	(0.51)	(0.22)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_0	49.2	248.2	497.18	2,483.55	4,968.35	24,837.09
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.044	0.054	0.059	0.083	0.117	0.390
	Y_1	Mean $\hat{\alpha}_1$	-0.03	-0.03	-0.04	-0.04	-0.04	-0.04
		Std. Dev. $\hat{\alpha}_1$	(0.50)	(0.23)	(0.15)	(0.07)	(0.05)	(0.02)
		Mean N_1	50.8	251.8	502.82	2,516.45	5,031.65	25,162.91
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.045	0.059	0.056	0.071	0.133	0.418
	Mean 1st Stage F-stats		0.99	1.25	1.34	3.02	4.91	21.15
0.15	Y_0	Mean $\hat{\alpha}_0$	0.10	0.12	0.13	0.12	0.11	0.12
		Std. Dev. $\hat{\alpha}_0$	(0.51)	(0.22)	(0.17)	(0.07)	(0.05)	(0.02)
		Mean N_0	49.08	244.83	490.6	2,448.87	4,899.37	24,505.76
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.061	0.079	0.134	0.377	0.632	0.999
	Y_1	Mean $\hat{\alpha}_1$	-0.11	-0.12	-0.12	-0.11	-0.11	-0.11
		Std. Dev. $\hat{\alpha}_1$	(0.51)	(0.23)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_1	50.92	255.17	509.4	2,551.13	5,100.63	25,494.24
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.055	0.094	0.120	0.354	0.594	1.000
	Mean 1st Stage F-stats		1.30	2.81	4.62	18.86	36.87	180.62

Table B1(b): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Different Sample Sizes (2/3)

γ^d			N =100	N =500	N =1,000	N =5,000	N =10,000	N =50,000
0.25	Y_0	Mean $\hat{\alpha}_0$	0.22	0.18	0.20	0.19	0.20	0.19
		Std. Dev. $\hat{\alpha}_0$	(0.50)	(0.22)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_0	48.15	241.11	482.33	2,416.13	4,834.96	24,168.62
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.063	0.122	0.216	0.756	0.975	1.000
	Y_1	Mean $\hat{\alpha}_1$	-0.17	-0.19	-0.19	-0.19	-0.19	-0.19
		Std. Dev. $\hat{\alpha}_1$	(0.48)	(0.22)	(0.15)	(0.07)	(0.05)	(0.02)
		Mean N_1	51.85	258.89	517.67	2,583.87	5,165.04	25,831.38
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.055	0.126	0.219	0.784	0.959	1.000
	Mean 1st Stage F-stats		2.10	6.00	11.31	51.28	102.22	497.11
0.5	Y_0	Mean $\hat{\alpha}_0$	0.40	0.39	0.39	0.39	0.39	0.39
		Std. Dev. $\hat{\alpha}_0$	(0.51)	(0.22)	(0.16)	(0.08)	(0.05)	(0.02)
		Mean N_0	47.06	234.2	468.45	2,340.41	4,676.3	23,387.59
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.096	0.388	0.652	1.000	1.000	1.000
	Y_1	Mean $\hat{\alpha}_1$	-0.35	-0.38	-0.38	-0.37	-0.37	-0.37
		Std. Dev. $\hat{\alpha}_1$	(0.49)	(0.21)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_1	52.94	265.8	531.55	2,659.59	5,323.7	26,612.41
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.108	0.391	0.636	1.000	1.000	1.000
	Mean 1st Stage F-stats		5.20	21.44	41.83	203.52	406.88	2,021.02

Table B1(c): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Different Sample Sizes (3/3)

γ^d			N =100	N =500	N =1,000	N =5,000	N =10,000	N =50,000
1.5	Y_0	Mean $\hat{\alpha}_0$	1.23	1.22	1.23	1.23	1.22	1.22
		Std. Dev. $\hat{\alpha}_0$	(0.69)	(0.29)	(0.21)	(0.09)	(0.07)	(0.03)
		Mean N_0	42.78	213.11	425.57	2,126.19	4,250	21,259.33
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.388	0.978	1.000	1.000	1.000	1.000
	Y_1	Mean $\hat{\alpha}_1$	-1.02	-1.01	-1.03	-1.02	-1.02	-1.02
		Std. Dev. $\hat{\alpha}_1$	(0.57)	(0.23)	(0.16)	(0.07)	(0.05)	(0.02)
		Mean N_1	57.22	286.89	574.43	2,873.81	5,750	28,740.67
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.452	0.991	1.000	1.000	1.000	1.000
	Mean 1st Stage F-stats		43.60	206.60	409.95	2,045.60	4,091.17	20,458.59
2.5	Y_0	Mean $\hat{\alpha}_0$	2.05	2.08	2.08	2.09	2.08	2.08
		Std. Dev. $\hat{\alpha}_0$	(1.19)	(0.52)	(0.36)	(0.16)	(0.11)	(0.05)
		Mean N_0	42.12	211.08	422.32	2,113.3	4,230.46	21,138.08
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.341	0.953	0.998	1.000	1.000	1.000
	Y_1	Mean $\hat{\alpha}_1$	-1.52	-1.54	-1.54	-1.55	-1.54	-1.55
		Std. Dev. $\hat{\alpha}_1$	(0.66)	(0.27)	(0.19)	(0.09)	(0.06)	(0.03)
		Mean N_1	57.88	288.92	577.68	2,886.7	5,769.54	28,861.92
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.616	1.000	1.000	1.000	1.000	1.000
	Mean 1st Stage F-stats		146.05	692.68	1,372.45	6,801.57	13,623.12	68,008.21

Notes. *: p¡0.05, **: p¡0.01, ***: p¡0.001. This table reports the estimation results for the first set of Monte Carlo analyses, where we fix the degree of selection on unobservables at $\rho = 0.6$. Y_0 and Y_1 refer to the dependent variable in the auxiliary regression. N refers to the sample size. γ^d represents the strength of the instrument in the data generating process. The "Mean $\hat{\alpha}_0$" and "Mean $\hat{\alpha}_1$" rows report the average coefficient estimates on the auxiliary regressor Z_i, which is our test statistic. The "Std. Dev. $\hat{\alpha}_0$" and "Std. Dev. $\hat{\alpha}_1$" rows report its standard deviations. The "Mean N_0" and "Mean N_1" rows provide the average number of $D_i = 0$ and $D_i = 1$ units, respectively. The "Prob. Rej. $H_0 : \alpha_0 = 0$" and "Prob. Rej. $H_0 : \alpha_1 = 0$" rows report the proportion of samples in which the data reject the corresponding null when the level of the test is 0.05. The "Mean 1st Stage F-stats" rows reports the average first-stage F-statistics. All reported means and standard deviations refer to the 1,000 MC samples.

Table B2(a): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Selection on Unobservables (N =1,000) (1/3)

γ^d			ρ =0	ρ =0.2	ρ =0.4	ρ =0.6
0	Y_0	Mean $\hat{\alpha}_0$	0.00	0.00	-0.01	0.00
		Std. Dev. $\hat{\alpha}_0$	(0.18)	(0.17)	(0.17)	(0.15)
		Mean N_0	499.98	499.71	499.21	499.16
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.057	0.037	0.046	0.040
	Y_1	Mean $\hat{\alpha}_1$	-0.01	0.00	0.00	0.00
		Std. Dev. $\hat{\alpha}_1$	(0.17)	(0.18)	(0.17)	(0.15)
		Mean N_1	500.02	500.29	500.79	500.84
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.044	0.056	0.039	0.050
	Mean 1st Stage F-stats		1.01	1.12	1.03	1.02
0.05	Y_0	Mean $\hat{\alpha}_0$	0.01	0.01	0.03	0.05
		Std. Dev. $\hat{\alpha}_0$	(0.18)	(0.18)	(0.17)	(0.16)
		Mean N_0	496.96	497.11	496.12	495.44
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.046	0.051	0.050	0.058
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.01	-0.03	-0.02
		Std. Dev. $\hat{\alpha}_1$	(0.18)	(0.18)	(0.17)	(0.16)
		Mean N_1	503.04	502.89	503.88	504.56
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.047	0.059	0.053	0.061
	Mean 1st Stage F-stats		1.34	1.33	1.50	1.39
0.15	Y_0	Mean $\hat{\alpha}_0$	0.00	0.04	0.07	0.12
		Std. Dev. $\hat{\alpha}_0$	(0.18)	(0.17)	(0.17)	(0.16)
		Mean N_0	489.97	489.57	490.26	490.06
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.049	0.042	0.061	0.103
	Y_1	Mean $\hat{\alpha}_1$	0.01	-0.04	-0.08	-0.11
		Std. Dev. $\hat{\alpha}_1$	(0.17)	(0.18)	(0.18)	(0.16)
		Mean N_1	510.03	510.43	509.74	509.94
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.044	0.056	0.096	0.104
	Mean 1st Stage F-stats		4.73	4.70	4.45	4.52

Table B2(b): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Selection on Unobservables (N =1,000) (2/3)

γ^d			ρ =0	ρ =0.2	ρ =0.4	ρ =0.6
0.25	Y_0	Mean $\hat{\alpha}_0$	-0.01	0.06	0.13	0.19
		Std. Dev. $\hat{\alpha}_0$	(0.19)	(0.19)	(0.18)	(0.17)
		Mean N_0	484.46	483.00	483.19	483.43
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.057	0.074	0.123	0.215
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.06	-0.13	-0.19
		Std. Dev. $\hat{\alpha}_1$	(0.17)	(0.18)	(0.17)	(0.15)
		Mean N_1	515.54	517.00	516.81	516.57
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.039	0.069	0.117	0.213
	Mean 1st Stage F-stats		10.97	10.87	10.82	11.32
0.5	Y_0	Mean $\hat{\alpha}_0$	0.00	0.14	0.26	0.40
		Std. Dev. $\hat{\alpha}_0$	(0.20)	(0.19)	(0.18)	(0.17)
		Mean N_0	467.71	467.45	467.4	467.73
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.053	0.116	0.289	0.654
	Y_1	Mean $\hat{\alpha}_1$	-0.01	-0.12	-0.25	-0.38
		Std. Dev. $\hat{\alpha}_1$	(0.18)	(0.18)	(0.17)	(0.15)
		Mean N_1	532.29	532.55	532.6	532.27
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.055	0.103	0.321	0.680
	Mean 1st Stage F-stats		42.06	40.94	41.36	42.61

Table B2(c): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Selection on Unobservables (N =1,000) (3/3)

γ^d			ρ =0	ρ =0.2	ρ =0.4	ρ =0.6
1.5	Y_0	Mean $\hat{\alpha}_0$	0.00	0.40	0.81	1.22
		Std. Dev. $\hat{\alpha}_0$	(0.24)	(0.25)	(0.24)	(0.21)
		Mean N_0	425.67	424.54	423.89	424.25
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.051	0.374	0.926	1.00
	Y_1	Mean $\hat{\alpha}_1$	-0.01	-0.35	-0.68	-1.02
		Std. Dev. $\hat{\alpha}_1$	(0.20)	(0.19)	(0.18)	(0.16)
		Mean N_1	574.33	575.46	576.11	575.75
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.061	0.469	0.959	1.00
	Mean 1st Stage F-stats		411.36	410.20	413.54	416.12
2.5	Y_0	Mean $\hat{\alpha}_0$	0.02	0.69	1.40	2.10
		Std. Dev. $\hat{\alpha}_0$	(0.45)	(0.42)	(0.40)	(0.36)
		Mean N_0	422.84	422.76	422.5	422.17
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.061	0.388	0.914	0.999
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.52	-1.04	-1.55
		Std. Dev. $\hat{\alpha}_1$	(0.22)	(0.22)	(0.21)	(0.19)
		Mean N_1	577.16	577.24	577.5	577.83
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.049	0.644	0.998	1.000
	Mean 1st Stage F-stats		1,373.49	1,376.46	1,373.56	1,366.90

Notes. *: p¡0.05, **: p¡0.01, ***: p¡0.001. This table reports the estimation results for the second set of Monte Carlo analyses, where we fix the sample size at $N = 1,000$. Y_0 and Y_1 refer to the dependent variable in the auxiliary regression. N refers to the sample size. γ^d represents the strength of the instrument in the data generating process. The "Mean $\hat{\alpha}_0$" and "Mean $\hat{\alpha}_1$" rows report the average coefficient estimates on the auxiliary regressor Z_i, which is our test statistic. The "Std. Dev. $\hat{\alpha}_0$" and "Std. Dev. $\hat{\alpha}_1$" rows report its standard deviations. The "Mean N_0" and "Mean N_1" rows provide the average number of $D_i = 0$ and $D_i = 1$ units, respectively. The "Prob. Rej. $H_0 : \alpha_0 = 0$" and "Prob. Rej. $H_0 : \alpha_1 = 0$" rows report the proportion of samples in which the data reject the corresponding null when the level of the test is 0.05. The "Mean 1st Stage F-stats" rows reports the average first-stage F-statistics. All reported means and standard deviations refer to the 1,000 MC samples.

Table B2(d): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Selection on Unobservables (N =10,000) (1/3)

γ^d			ρ =0	ρ =0.2	ρ =0.4	ρ =0.6
0	Y_0	Mean $\hat{\alpha}_0$	0.00	0.00	0.01	0.00
		Std. Dev. $\hat{\alpha}_0$	(0.06)	(0.06)	(0.05)	(0.05)
		Mean N_0	4,997.33	4,997.92	4,999.98	4,997.98
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.056	0.049	0.052	0.052
	Y_1	Mean $\hat{\alpha}_1$	0.00	0.00	0.00	0.00
		Std. Dev. $\hat{\alpha}_1$	(0.05)	(0.05)	(0.05)	(0.05)
		Mean N_1	5,002.67	5,002.08	5,000.02	5,002.02
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.054	0.042	0.046	0.041
	Mean 1st Stage F-stats		1.02	0.99	1.00	1.01
0.05	Y_0	Mean $\hat{\alpha}_0$	0.00	0.01	0.02	0.04
		Std. Dev. $\hat{\alpha}_0$	(0.06)	(0.06)	(0.05)	(0.05)
		Mean N_0	4,964.93	4,968.46	4,969.12	4,965.2
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.057	0.055	0.069	0.113
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.01	-0.03	-0.04
		Std. Dev. $\hat{\alpha}_1$	(0.06)	(0.06)	(0.05)	(0.05)
		Mean N_1	5,035.07	5,031.54	5,030.88	5,034.8
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.047	0.047	0.088	0.118
	Mean 1st Stage F-stats		4.83	5.04	4.81	4.79
0.15	Y_0	Mean $\hat{\alpha}_0$	0.00	0.04	0.08	0.12
		Std. Dev. $\hat{\alpha}_0$	(0.06)	(0.06)	(0.05)	(0.05)
		Mean N_0	4,898.89	4,901.47	4,899.88	4,902.14
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.046	0.115	0.328	0.663
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.04	-0.08	-0.12
		Std. Dev. $\hat{\alpha}_1$	(0.06)	(0.05)	(0.05)	(0.05)
		Mean N_1	5,101.11	5,098.53	5,100.12	5,097.86
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.051	0.103	0.295	0.635
	Mean 1st Stage F-stats		36.95	36.34	37.39	36.81

Table B2(e): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Selection on Unobservables (N =10,000) (2/3)

γ^d			ρ =0	ρ =0.2	ρ =0.4	ρ =0.6
0.25	Y_0	Mean $\hat{\alpha}_0$	0.00	0.06	0.13	0.20
		Std. Dev. $\hat{\alpha}_0$	(0.06)	(0.06)	(0.05)	(0.05)
		Mean N_0	4,834.28	4,834.52	4,834.20	4,835.94
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.047	0.210	0.657	0.975
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.06	-0.13	-0.19
		Std. Dev. $\hat{\alpha}_1$	(0.06)	(0.05)	(0.05)	(0.05)
		Mean N_1	5,165.72	5,165.48	5,165.80	5,164.06
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.051	0.189	0.666	0.962
	Mean 1st Stage F-stats		102.23	101.65	101.06	100.94
0.5	Y_0	Mean $\hat{\alpha}_0$	0.00	0.13	0.26	0.39
		Std. Dev. $\hat{\alpha}_0$	(0.06)	(0.06)	(0.06)	(0.05)
		Mean N_0	4,678.93	4,678.31	4,677.16	4,678.46
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.055	0.610	0.997	1.00
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.12	-0.25	-0.37
		Std. Dev. $\hat{\alpha}_1$	(0.06)	(0.06)	(0.05)	(0.05)
		Mean N_1	5,321.07	5,321.69	5,322.84	5,321.54
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.052	0.620	0.998	1.000
	Mean 1st Stage F-stats		405.87	403.92	405.56	407.14

Table B2(f): Monte Carlo Analyses on Varying Degree of Instrument Strengths and Selection on Unobservables (N =10,000) (3/3)

γ^d			ρ =0	ρ =0.2	ρ =0.4	ρ =0.6
1.5	Y_0	Mean $\hat{\alpha}_0$	0.00	0.41	0.81	1.22
		Std. Dev. $\hat{\alpha}_0$	(0.08)	(0.07)	(0.07)	(0.07)
		Mean N_0	4,251.75	4,253.15	4,248.74	4,251.16
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.058	0.999	1.000	1.000
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.34	-0.68	-1.02
		Std. Dev. $\hat{\alpha}_1$	(0.06)	(0.06)	(0.06)	(0.05)
		Mean N_1	5,748.25	5,746.85	5,751.26	5,748.84
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.038	1.000	1.000	1.000
		Mean 1st Stage F-stats	4,104.46	4,091.92	4,093.64	4,098.88
2.5	Y_0	Mean $\hat{\alpha}_0$	0.00	0.69	1.39	2.07
		Std. Dev. $\hat{\alpha}_0$	(0.13)	(0.13)	(0.12)	(0.11)
		Mean N_0	4,228.44	4,226.26	4,229.89	4,226.3
		Prob. Rej. $H_0 : \alpha_0 = 0$	0.048	0.999	1.000	1.000
	Y_1	Mean $\hat{\alpha}_1$	0.00	-0.52	-1.03	-1.55
		Std. Dev. $\hat{\alpha}_1$	(0.07)	(0.07)	(0.06)	(0.06)
		Mean N_1	5,771.56	5,773.74	5,770.11	5,773.7
		Prob. Rej. $H_0 : \alpha_1 = 0$	0.045	1.000	1.000	1.000
		Mean 1st Stage F-stats	13,607.83	13,601.76	13,616.54	13,583.13

Notes. *: p¡0.05, **: p¡0.01, ***: p¡0.001. This table reports the estimation results for the second set of Monte Carlo analyses, where we fix the sample size at $N = 10,000$. Y_0 and Y_1 refer to the dependent variable in the auxiliary regression. N refers to the sample size. γ^d represents the strength of the instrument in the data generating process. The "Mean $\hat{\alpha}_0$" and "Mean $\hat{\alpha}_1$" rows report the average coefficient estimates on the auxiliary regressor Z_i, which is our test statistic. The "Std. Dev. $\hat{\alpha}_0$" and "Std. Dev. $\hat{\alpha}_1$" rows report its standard deviations. The "Mean N_0" and "Mean N_1" rows provide the average number of $D_i = 0$ and $D_i = 1$ units, respectively. The "Prob. Rej. $H_0 : \alpha_0 = 0$" and "Prob. Rej. $H_0 : \alpha_1 = 0$" rows report the proportion of samples in which the data reject the corresponding null when the level of the test is 0.05. The "Mean 1st Stage F-stats" rows reports the average first-stage F-statistics. All reported means and standard deviations refer to the 1,000 MC samples.