DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

On the Practice of Lagging Variables To Avoid Simultaneity

W. Robert Reed

WORKING PAPER

No. 18/2014

Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 18/2014

On the Practice of Lagging Variables To Avoid Simultaneity

W. Robert Reed[1]

July 6, 2014

Abstract: A common practice in applied economics research consists of replacing a suspected simultaneously-determined explanatory variable with its lagged value. This note demonstrates that this practice does not enable one to avoid simultaneity bias. The associated estimates are still inconsistent, and hypothesis testing is invalid. One alternative is to use lagged values of the endogenous variable in instrumental variable estimation. However, this is only an effective estimation strategy if the lagged values do not themselves belong in the respective estimating equation, and if they are sufficiently correlated with the simultaneously-determined explanatory variable.

NOTE: This paper is a revision of Working Paper No. 13/32.

Keywords: Simultaneity, Reverse causality, Lagged variables

JEL Classifications: C1, C5, C15

Acknowledgements: I acknowledge helpful comments from Kuntal Das, Chris Hajzler, Steven Stillman, Jeffrey Wooldridge, James Ziliak, Arthur Grimes, and seminar participants at the University of Otago and the 2014 New Zealand Association of Economists conference. Remaining errors are my own.

[1] Department of Economics and Finance, University of Canterbury, New Zealand

*Corresponding Author: bob.reed@canterbury.ac.nz

I. INTRODUCTION

Simultaneity is a concern in much of empirical economic analysis. One approach that has been employed to avoid the problems associated with simultaneity is to replace the suspect explanatory variable with its lagged value. The practice is widespread, as can be confirmed by searching for variations of "avoid simultaneity lagged variables" on Google Scholar. Recent examples are Aschoff and Schmidt (2008); Bania, Gray, and Stone (2007); Bansak, Morin and Starr (2007); Brinks and Coppedge (2006); Buch, Koch, and Koetter (2013); Clemens, Radelet, Bhavnani and Bazzi (2012); Cornett, Marcus, Saunders, and Tehranian (2007); Green Malpezzi, and Mayo (2005); Gupta (2005); Hayo, Kutan, and Neuenkirch (2010); Jensen and Paldam (2006), MacKay and Phillips (2005); Spilimbergo (2009); Stiebala (2011); and Vergara (2010). The practice is common across a wide variety of disciplines in economics and finance. Many appear in top journals including the *American Economic Review*, the *Journal of Finance,* the *Economic Journal*, and the *Journal of Banking & Finance*, and are highly cited.

The rationale for the practice is explicitly identified in statements such as the following: "We avoid poor-quality instrumental variables and instead address potential biases from reverse and simultaneous causation by … lagging" (Clemens, Radelet, Bhavnani and Bazzi, 2012); "The vector of controls contains lagged returns…Contemporaneous U.S. returns are excluded to avoid simultaneity problems" (Hayo, Kutan, and Neuenkirch, 2010); and "The variable is expressed as a percentage of GDP. The lagged variable was used in both cases to avoid possible simultaneity problems" (Vergara, 2010).[1]

[1] Other examples are: "Innovation intensity is included as a lagged variable in order to mitigate simultaneity problems" (Aschhoff and Schmidt, 2008, p. 48); "The variables ΔIP, *I/K*, and STDEV are intended to capture effects on utilization of output growth, investment level, and output volatility, respectively; they are included in lagged form to avoid problems with simultaneity" (Bansak et al., 2007, p. 636f.); "…so long as we avoid the simultaneity problem and incorporate the main sources of common shocks to the countries at issue (as we do by lagging the diffusion variable and including the most important domestic variables), our estimates should be relatively unbiased and consistent" (Brinks and Coppedge, 2006, p. 476); "We lag the explanatory variables

The purpose of this note is to draw attention to the fact that replacing a contemporaneous explanatory variable with its lagged value does not avoid the inconsistency problems associated with simultaneity. In contrast, using lagged values of the endogeneous explanatory variable and/or dependent variable as instruments can provide an effective estimation strategy if (i) the lagged values do not themselves belong in the respective estimating equation, and (ii) they are sufficiently correlated with the simultaneously-determined explanatory variable.

II. THEORY

Let Y be a function of either, or both, contemporaneous and lagged X and let us assume that the effect of X on Y is represented by the following relationship:

(1) $$Y_t = a + bX_t + cX_{t-1} + \varepsilon_t,$$

where $\varepsilon_t \sim NID(0, \sigma_Y)$, and b and/or c may be zero. A researcher suspects that Y and X are simultaneously determined. In an effort to avoid simultaneity bias, the researcher estimates

(1') $$Y_t = \alpha + \beta X_{t-1} + error\ term.$$

To determine the relationship between β and b and c, one needs to know how X is affected by contemporaneous Y and (possibly) lagged X. Let us assume this relationship is represented by

(2) $$X_t = d + eY_t + fX_{t-1} + \nu_t,$$ where $\nu_t \sim NID(0, \sigma_X)$,

---

$X_{i,t-1}$ by one period to avoid simultaneity" (Buch et al., 2013, p. 1415); "We lag all measures of institutional ownership and institutional board membership by one year. This lag allows for the effect of any change in governance structure to show up in firm performance. This also mitigates simultaneity issues" (Cornett et al., 2007, p. 1781); "We therefore also perform regressions with lagged changes to avoid simultaneity problems" (Green et al., 2005, p. 335f.); "First, to minimize the possibility of simultaneity between privatization and performance, we investigate the impact of the lagged share of private ownership on current performance" (Gupta, 2005, p. 989); ""…we cannot a priori reject reverse causality. Hence, we need to control for countercausality in aid-growth regressions. Three methods are available: (1) Aid is lagged by one time unit relative to the growth explained..."; "We lag the industry medians to avoid endogeneity problems" (MacKay and Phillips, 2005, p. 1450); "To avoid simultaneity bias, this specification has explanatory variables lagged five years in the five-year specifications as well" (Spilimbergo, 2009, p. 536); "In all specifications lagged values of the financial indicators are used. This is to allow for a time lag between financial development and the export decision, since planning and realisation of foreign market entry and expansion might take time. The use of lagged values also reduces simultaneity problems" (Stiebale, 2011, p. 130).

Then

(3) $$Y_t = \frac{(a+bd)}{(1-be)} + \frac{(bf+c)}{(1-be)} X_{t-1} + \frac{b}{(1-be)} \nu_t + \frac{1}{(1-be)} \varepsilon_t \,.$$ [2]

It follows that OLS estimation of Equation (1') produces a consistent estimate of the reduced form coefficient on lagged *X*,

(4) $$plim(\hat{\beta}) = \frac{(bf+c)}{(1-be)},$$

where simultaneity is represented by the parameter *e*, and serial correlation in *X* by *f*. Equation (4) makes clear that it is not generally possible to recover the structural parameters *b* and *c* from $\hat{\beta}$.

A similar problem arises if one regresses the change in *Y* on lagged *X*. In this case,

(5) $$\Delta Y_t = \frac{(a+bd)}{(1\text{-}be)} + \frac{(c+bf)}{(1-be)} X_{t-1} - Y_{t-1} + \frac{b}{(1-be)} \nu_t + \frac{1}{(1-be)} \varepsilon_t \,.$$

OLS estimation of

(5') $$\Delta Y_t = \alpha + \beta X_{t-1} + \gamma Y_{t-1} + \ error\ term,$$

produces a consistent estimate of the reduced form coefficient on lagged *X*, so that again

$$plim(\hat{\beta}) = \frac{(bf+c)}{(1-be)},$$

and recovery of *b* and *c* is not generally possible.

FIGURE 1 illustrates the problem facing the researcher interested in estimating the structural parameters *b* and/or *c*. For given values of *b* and *c*, the sign, size, and statistical significance of $\hat{\beta}$ is greatly influenced by the size of the simultaneity parameter, *e*, even though contemporaneous *X* is excluded from the estimating equation. For sufficiently large *e*, the estimated sign of $\hat{\beta}$ can be the opposite of *b* and/or *c*. The figure illustrates the case where *b=1*, *c=0*, and *f=0.5*. Given these values, any value of *e > 1* produces $plim(\hat{\beta})$<*0*. I

[2] The corresponding value of X_t is $X_t = \frac{(d+ae)}{(1-be)} + \frac{(f+ce)}{(1-be)} X_{t-1} + \frac{1}{(1-be)} \nu_t + \frac{e}{(1-be)} \varepsilon_t$. Accordingly, $\frac{(f+ce)}{(1-be)} < 1$ is necessary if X_t is to avoid explosive dynamic behaviour.

next consider three cases and identify sufficient conditions for the researcher to identify the desired structural parameter(s).

CASE ONE: Estimation of *b*. This first case represents the scenario where a researcher replaces current *X* with its lagged value, and assumes that the coefficient on lagged *X* tells him/her something about the direct effect of current *X* on current *Y*. From Equation (4), it follows that sufficient conditions for $\hat{\beta}$ to consistently estimate *b* are (i) $c=0$, (ii) $e=0$, and (iii) $f=1$.[3] The first condition states that only current *X*, not lagged *X*, exerts a direct effect on *Y*. The second condition states that *X* is not simultaneously determined, so there is no simultaneity problem. The third condition is that *X*, and thus *Y*, are random walk processes. When all three conditions hold, OLS estimates *b* consistently.[4] Of course, if there is no simultaneity problem, there is no cause to replace X_t with X_{t-1}. The researcher should regress *Y* directly on current *X*.

CASE TWO: Estimation of the "total direct effect," $b+c$. The second case represents the scenario where the researcher believes that both current and lagged *X* directly affect *Y*, and that the coefficient on lagged *X* allows him/her to estimate the "total effect" ($b+c$) free from simultaneity bias. This case is similar to the previous case. Sufficient conditions for $\hat{\beta}$ to consistently estimate the sum ($b+c$) are (i) $e=0$ and (ii) $f=1$. In other words, the strategy of substituting lagged *X* for current *X* will produce consistent estimates of ($b+c$) when there is no simultaneity and *X* is a random walk. As before, if these conditions hold, there is no reason to replace X_t with X_{t-1}. The researcher should regress *Y* directly on current *X* to obtain a consistent estimate of ($b+c$).

CASE THREE: Estimation of *c*. The third case represents the scenario where the researcher believes *X* affects *Y* with a lag. Two sets of conditions are noteworthy. OLS will produce consistent estimates of *c* if (i) $e=0$, and (ii) $f=0$. Under these conditions, there is no

[3] An alternative set of "knife-edge" conditions can be obtained by setting $\frac{(bf+c)}{(1-be)} = b$.

[4] In this case, Y_t and X_t will be cointegrated and OLS estimates will be superconsistent.

simultaneity bias and no omitted variable bias from omitting X_t from the estimating equation. The specification used for the estimating equation correctly specifies the DGP. A second regime by which least squares regression produces a consistent estimate of c occurs when $b=0$; i.e., where there is again no simultaneity bias and no omitted variable bias from excluding X_t from the estimating equation.

The three cases above illustrate the ineffectiveness of the practice of lagging explanatory variables to avoid simultaneity bias. A common denominator in all three cases is that the practice of replacing X_t with X_{t-1} produces consistent estimates when there is no simultaneity; in which case the researcher should just use X_t.

III. SIMULATION RESULTS

The preceding section demonstrates that lagging X does not enable one to escape simultaneity bias. In this section, I use simulations to illustrate how the data can mislead the researcher into making incorrect inferences about the true the effect of X on Y.

TABLE 1 presents three sets of simulation results. In the first two sets of simulations, the DGP is:

6a) $$\begin{aligned} Y_t &= 1\, X_t + 0\, X_{t-1} + \varepsilon_t \\ X_t &= 5\, Y_t + f\, X_{t-1} + v_t \end{aligned} \quad , \ \varepsilon_{t,}, v_t \sim NID(0,1);$$

where f is alternatively set equal to 0 and 0.5; Y_t and X_t are simultaneously determined, and the true, direct effect of X_{t-1} on Y_t is zero ($c=0$). SIMULATION 1 (SIMULATION 2) simulates a DGP where X is not (is) characterized by serial correlation. These simulations illustrate how serial correlation in X_t can substantially impact intepretation of the empirical results from regressing Y_t on X_{t-1}.

For each set of simulations I use OLS to estimate the equation

6b) $Y_t = \alpha + \beta\, X_{t-1} + error\ term.$

The simulated datasets vary in size from $T=10$ to $T=1000$. For each value of T, 10,000 data sets are generated, producing 10,000 estimates of β, the coefficient on X_{t-1} in Equation (6b). The table reports the mean estimate of β for each set of replications, along with the rate at which the null hypothesis, $H_0\colon \beta = 0$, is rejected.

In SIMULATION 1, $b=1$, $c=0$, $e=5$, and $f=0$, so that $plim(\hat{\beta}) = \frac{(bf+c)}{(1-be)} = 0$. The mean estimated value of β suffers from finite sample bias, but converges to its probability limit as the sample sizes increase. The rejection rates for $H_0\colon \beta = 0$ are close to 0.05. If the researcher believes that $\hat{\beta}$ is a measure of the effect of X on Y, he/she will incorrectly conclude that X has no effect on Y approximately 95% of the time, despite the fact that the true effect of X on Y is 1.

SIMULATION 2 shows how serial correlation can alter the estimated relationship. Everything is identical to the first set of simulations except that $f=0.5$. Accordingly, $plim(\hat{\beta}) = \frac{(bf+c)}{(1-be)} = -0.125$. As before, the mean estimated value of β suffers from finite sample bias, but converges to its asymptotic value as the sample sizes increase. When $T=10$, approximately a fourth of all regressions result in rejection of the null. This rises to half when $T=20$. By the time $T=50$, almost 90 percent of regressions produce a rejection of the null hypothesis. Accordingly, a researcher who thinks he/she is estimating the effect of X on Y will incorrectly conclude in the vast majority of cases that X is negatively and significantly associated with Y, even though the true, direct effect of X on Y is positive and equal to 1.

SIMULATION 3 repeats the analysis of SIMULATION 2 except that the dependent variable in the DGP is ΔY_t rather than Y_t, and the estimated equation is now $\Delta Y_t = \alpha + \beta X_{t-1} + \gamma Y_{t-1} + \textit{error term}$. Comparison of Equation (5) with Equation (3) suggests that the results should be similar to SIMULATION 2, and indeed this is the case. The estimates of β converge to their reduced form value of $\frac{(c+bf)}{(1-be)} = -0.125$, and the rejection rate of the null converges to 1.000 as the sample size increases, albeit at a somewhat slower rate than in the level case. While not reported here, I have also simulated panel data sets with the same parameter values as above and obtained similar results.[5]

IV. A CONSISTENT ESTIMATION STRATEGY

Equations (1) and (2) comprise a two-equation, endogenous system in variables X_t, Y_t, and X_{t-1}. When the lagged X variable does not appear in Equation (1) -- so that c=0 -- X_{t-1} and/or Y_{t-1} can serve as valid instruments for X_t in Equation (1), assuming that these are correlated with the endogenous X_t.

TABLE 2 reports three sets of simulations corresponding to the DGP:

$$7)\quad \begin{aligned} Y_t &= 1\, X_t + 0\, X_{t-1} + \varepsilon_t \\ X_t &= 5\, Y_t + f\, X_{t-1} + v_t \end{aligned} \qquad , \ \varepsilon_{t,}, v_t \sim NID(0,1);$$

where f takes values 0, 0.5, and 0.9. When $f = 0$, X_t and X_{t-1} are uncorrelated. When $f = 0.5$ and 0.9, the correlation between X_t and X_{t-1} is -0.125 and -0.225, respectively. Note that the other parameters are specified to match the values of the first two simulations in TABLE 1, so that the results can be compared with the corresponding simulations from that table.

For each set of simulations, I use three different pairs of instruments: (i) X_{t-1} and X_{t-2}, (ii) X_{t-1} and Y_{t-1}, and (iii) Y_{t-1} and Y_{t-2}.[6] The simulations in TABLE 2 use 2SLS to estimate the model $Y_t = \alpha + \beta\, X_t + \textit{error term}$. In Panel A, X_t and X_{t-1} are uncorrelated, so that 2SLS using the lagged values as instruments is not consistent. In contrast, in Panels B and C, when $f = 0.5$ and 0.9, the mean value of the 2SLS estimator is very close to its asymptotic value given 1000 observations. In smaller samples, the 2SLS estimator remains substantially biased, though the bias gets smaller as the correlation of X_t and X_{t-1} increases

[5] The associated results, as well as Stata .do files for all simulations, are available from the author upon request.

[6] I thank a reviewer for suggesting these instruments. Note that when the degree of overidentification equals 1, as it does in this case, the mean of the 2SLS estimator exists, but its variance does not (Kinal, 1980). I thank David Giles for pointing this out to me.

(i.e., as f increases from 0.5 to 0.9). This illustrates the importance of having X_t and X_{t-1} being highly correlated.

When c $\neq$ 0, so that the lagged value of X appears in Equation (1), it is necessary to use deeper lags of X and Y as instruments for X_{t-1}. As suggested by TABLE 2, this will only be an effective strategy if these lags are sufficiently correlated with X_{t-1}.

V. CONCLUSION

A common practice in applied econometrics work consists of replacing a (suspected) simultaneously-determined explanatory variable with its lagged value. This note demonstrates that this practice does not enable one to escape simultaneity bias. I show through both theory and simulations the infeasibility of identifying structural parameters of the DGP when the relationship between X and Y is characterized by simultaneity. Further, I demonstrate that it is straightforward to generate examples where the researcher is likely to conclude that the effect of X on Y is opposite in sign to its true value, and to find that the associated, wrong-signed coefficient is statistically significant a majority of the time.

An alternative to the practice of substituting lagged values for contemporaneous variables, is to use the lagged values as instruments in 2SLS/GMM/LIML estimation. However, this is only an effective estimation strategy if the lagged values do not themselves belong in the respective estimating equation, and if they are sufficiently correlated with the simultaneously-determined explanatory variable. In any case, the implication of this study is that researchers should avoid the practice of lagging variables to circumvent the problems of simultaneity.

REFERENCES

Arellano, M., and S. Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277‑297.

Aschhoff, B., and T. Schmidt. 2008. Empirical evidence on the success of R&D cooperation—Happy together? *Review of Industrial Organization* 33:41–62

Bania, N., Gray, J.A., and J.A. Stone. 2007. Growth, taxes, and government expenditures: Growth hills for U.S. states. *National Tax Journal* LX: 193-204.

Bansak, C., Morin, N., and M. Starr. 2007. Technology, capital spending, and capacity utilization. *Economic Inquiry* 45:631–645

Brinks, D. and M. Coppedge. 2006. Diffusion is no illusion: Neighbor emulation in the third wave of democracy. *Comparative Political Studies* 39:463-489

Buch, C., Koch, C.T., and M. Koetter. 2013. Do banks benefit from internationalization? Revisiting the market power–risk nexus. *Review of Finance* 17:1401–1435.

Clemens, M.A., Radelet, S., Bhavnani, R.R., and S. Bazzi. 2012. Counting chickens when they hatch: Timing and the effects of aid on growth. *The Economic Journal* 122:590–617.

Cornett, M.M., Marcus, A.J., Saunders, A., and H. Tehranian. 2007. The impact of institutional ownership on corporate operating performance. *Journal of Banking & Finance* 31:1771–1794.

Green, R.K., Malpezzi, S., and S.K. Mayo. 2005. Metropolitan-Specific Estimates of the Price Elasticity of Supply of Housing, and Their Sources. *American Economic Review* 95:334-339.

Gupta. N. 2005. Partial privatization and firm performance. *Journal of Finance* 60:987-1015.

Hayo, B., Kutan, M.K., and M. Neuenkirch. 2010. The impact of U.S. central bank communication on European and Pacific equity markets. *Economics Letters* 108:172–174.

Jensen, P.S., and M. Paldam. 2006. Can the two new aid-growth models be replicated? *Public Choice* 127:147-175.

Kinal, T. 1980. The existence of moments of k-class estimators. *Econometrica* 48: 241-249.

MacKay, P. and G.M. Phillips. 2005. How does industry affect firm financial structure? *Review of Financial Studies* 18:1433-1466.

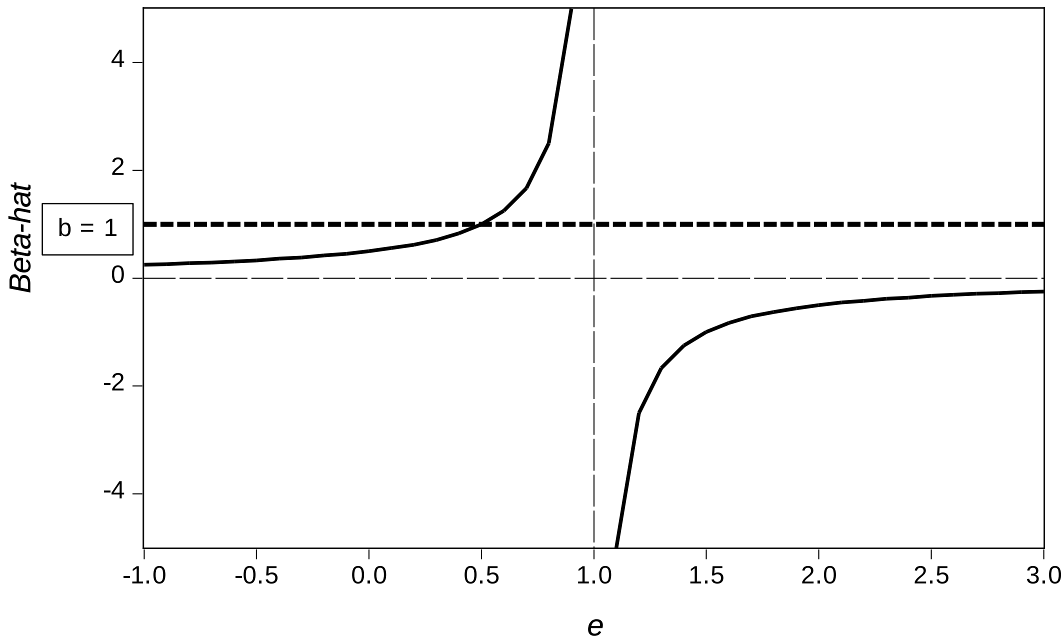
Nickell, S. J. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.

Spilimbergo, A. 2009. Democracy and Foreign Education. *American Economic Review* 99:528-543.

Stiebale, J. 2011. Do financial constraints matter for foreign market entry? A firm-level examination. *World Economy* 34:123–153.

Vergara, R. 2010. Taxation and private investment: Evidence for Chile. *Applied Economics*, 42:717-725.

FIGURE 1
The Relationship between $\widehat{\boldsymbol{\beta}}$ and the Simultaneity Parameter *e*



b = 1, c = 0, f = 0.5

NOTE #1: The parameters *b*, *c*, *e*, and *f* are identified in Equations (1) and (2) in the text. In the graph above, *b*, *c*, and *f* are fixed at 1, 0, and 0.5, respectively. The graph shows the relationship between $plim(\hat{\beta}) = \frac{(bf+c)}{(1-be)}$ and the simultaneity parameter, *e*. As a point of comparison, the total direct effect of *X* on *Y* is given by *b*+*c*. The graph illustrates the infeasibility of recovering the structural parameters *b* and *c* from the estimated coefficient on lagged *X* in the presence of simultaneity and serial correlation in *X*.

NOTE #2: A necessary condition for X_t to not possess explosive dynamic behaviour is $\frac{(f+ce)}{(1-be)}$ < 1. In the numerical example represented by the figure, this requires |e-1| > 0.5.

TABLE 1
Simulation Results: Addressing Simultaneity by Lagging X

A) SIMULATION 1: *Simultaneity but no Serial Correlation in X_t (Dep. Var. in Level Form)*

DGP: $1)\ Y_t = 1\,X_t + 0\,X_{t-1} + \varepsilon_t$, $\varepsilon_t, v_t \sim NID(0,1)$
$2)\ X_t = 5\,Y_t + 0\,X_{t-1} + v_t$

Estimated Equation: $Y_t = \alpha + \beta X_{t-1} + \text{error term}$

	T=10	T=20	T=50	T=100	T=1000
Mean $\hat{\beta}$	-0.0219	-0.0114	-0.0041	-0.0019	-0.0001
Rejection Rate for H_0: $\beta = 0$	0.039	0.043	0.046	0.051	0.049

B) SIMULATION 2: *Simultaneity with Serial Correlation in X_t (Dep. Var. in Level Form)*

DGP: $1)\ Y_t = 1\,X_t + 0\,X_{t-1} + \varepsilon_t$, $\varepsilon_t, v_t \sim NID(0,1)$
$2)\ X_t = 5\,Y_t + 0.5\,X_{t-1} + v_t$

Estimated Equation: $Y_t = \alpha + \beta X_{t-1} + \text{error term}$

	T=10	T=20	T=50	T=100	T=1000
Mean $\hat{\beta}$	-0.1397	-0.1325	-0.1274	-0.1260	-0.1251
Rejection Rate for H_0: $\beta = 0$	0.241	0.496	0.880	0.993	1.000

C) SIMULATION 3: *Simultaneity with Serial Correlation in X_t (Dep. Var. in Difference Form)*

DGP: $1)\ \Delta Y_t = 1\,X_t + 0\,X_{t-1} - Y_{t-1} + \varepsilon_t$, $\varepsilon_t, v_t \sim NID(0,1)$
$2)\ X_t = 5\,Y_t + 0.5\,X_{t-1} + v_t$

Estimated Equation: $\Delta Y_t = \alpha + \beta X_{t-1} + \gamma\, Y_{t-1} + \text{error term}$

	T=10	T=20	T=50	T=100	T=1000
Mean $\hat{\beta}$	-0.1568	-0.1411	-0.1317	-0.1283	-0.1251
Rejection Rate for H_0: $\beta = 0$	0.125	0.221	0.487	0.774	1.000

TABLE 2
Simulation Results: Addressing Simultaneity by Using Lagged X and Y as Instruments

	T=10	T=20	T=50	T=100	T=1000
A) DGP: $Y_t = 1\,X_t + 0\,X_{t-1} + \varepsilon_t$, $X_t = 5\,Y_t + 0\,X_{t-1} + v_t$, $\varepsilon_t, v_t \sim NID(0,1)$					
Instruments: X_{t-1}, X_{t-2}					
Mean $\hat{\beta}$	0.2426	0.2338	0.2370	0.2327	0.2275
Rejection Rate for H_0: $\beta = 0$	0.414	0.390	0.369	0.367	0.355
Instruments: X_{t-1}, Y_{t-1}					
Mean $\hat{\beta}$	0.2273	0.2334	0.2292	0.2263	0.2288
Rejection Rate for H_0: $\beta = 0$	0.421	0.394	0.367	0.365	0.352
Instruments: Y_{t-1}, Y_{t-2}					
Mean $\hat{\beta}$	0.2192	0.2275	0.2234	0.2300	0.2348
Rejection Rate for H_0: $\beta = 0$	0.410	0.389	0.366	0.357	0.355
B) DGP: $Y_t = 1\,X_t + 0\,X_{t-1} + \varepsilon_t$, $X_t = 5\,Y_t + 0.5\,X_{t-1} + v_t$, $\varepsilon_t, v_t \sim NID(0,1)$					
Instruments: X_{t-1}, X_{t-2}					
Mean $\hat{\beta}$	0.3287	0.3754	0.4968	0.6583	0.9938
Rejection Rate for H_0: $\beta = 0$	0.487	0.487	0.535	0.613	0.992
Instruments: X_{t-1}, Y_{t-1}					
Mean $\hat{\beta}$	0.3234	0.3782	0.4995	0.6553	0.9933
Rejection Rate for H_0: $\beta = 0$	0.496	0.492	0.539	0.610	0.993
Instruments: Y_{t-1}, Y_{t-2}					
Mean $\hat{\beta}$	0.2795	0.3344	0.4223	0.5780	0.9927
Rejection Rate for H_0: $\beta = 0$	0.446	0.437	0.483	0.543	0.967

C) DGP: $Y_t = 1\,X_t + 0\,X_{t-1} + \varepsilon_t$, $X_t = 5\,Y_t + 0.9\,X_{t-1} + \nu_t$, $\varepsilon_t, \nu_t \sim NID(0,1)$					
	T=10	T=20	T=50	T=100	T=1000
<u>Instruments</u>: X_{t-1}, X_{t-2}					
Mean $\widehat{\beta}$	0.4478	0.5584	0.7792	0.9434	0.9980
Rejection Rate for H_0: $\beta = 0$	0.530	0.570	0.710	0.846	1.000
<u>Instruments</u>: X_{t-1}, Y_{t-1}					
Mean $\widehat{\beta}$	0.4438	0.5625	0.7777	0.9215	0.9978
Rejection Rate for H_0: $\beta = 0$	0.533	0.573	0.705	0.847	1.000
<u>Instruments</u>: Y_{t-1}, Y_{t-2}					
Mean $\widehat{\beta}$	0.3725	0.4705	0.6737	0.8567	0.9999
Rejection Rate for H_0: $\beta = 0$	0.446	0.482	0.603	0.728	1.000

NOTE: In each of the simulations above, the estimated equation is $Y_t = \alpha + \beta X_t +$ *error term.* 2SLS is used to estimate β, with the instruments being, alternatively, (i) X_{t-1}, X_{t-2}; (ii) X_{t-1}, Y_{t-1}; and (iii) Y_{t-1}, Y_{t-2}. The three sets of simulations reported in panels A), B) and C) differ only in that the serial correlation parameter, f, takes values 0, 0.5, and 0.9, respectively.