DEPARTMENT OF ECONOMICS AND FINANCE

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

A NOTE ON THE PRACTICE OF LAGGING VARIABLES TO AVOID SIMULTANEITY

W. Robert Reed

WORKING PAPER

No. 32/2013

\

ISSN 1179-3228
Department of Economics and Finance
College of Business and Economics
University of Canterbury
Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 32/2013

A NOTE ON THE PRACTICE OF LAGGING VARIABLES TO AVOID SIMULTANEITY

W. Robert Reed[1]

6 October 2013

Abstract: A common practice in applied econometrics consists of replacing a suspected endogenous variable with its lagged values. This note demonstrates that lagging an endogenous variable does not enable one to escape simultaneity bias. The associated estimates are still inconsistent, and hypothesis testing is invalid. I show that the only time a strategy of replacing X_t with X_{t-1} enables consistent estimation of structural parameters is when there is no simultaneity to begin with. The implication of this study is that researchers who employ this practice should be explicit about why lagging is an effective estimation strategy.

Keywords: Simultaneity, Endogeneity, Lagged variables

JEL Classifications: C1, C5, C15

Acknowledgements: I acknowledge helpful comments from Kuntal Das, Chris Hajzler, Steven Stillman, Jeffrey Wooldridge, and seminar participants at the University of Otago. Remaining errors are my own.

1. Department of Economics and Finance, University of Canterbury, Christchurch, New Zealand.

*Corresponding Author: W. Robert Reed, Private Bag 4800, Christchurch 8140, New Zealand. Phone: +64-3-364-2846; Email: bob.reed@canterbury.ac.nz .

WORKING PAPER No. 32/2013

A NOTE ON THE PRACTICE OF LAGGING VARIABLES TO AVOID SIMULTANEITY

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 a suspected endogenous variable with its lagged values. 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); MacKay and Phillips (2005); Spilimbergo (2009); Stiebala (2011); and Vergara (2010). Many of these papers appear in top journals such as the *American Economic Review*, the *Journal of Finance*, the *Economic Journal*, and the *Journal of Banking & Finance.* Several 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

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. Some of the papers cited above further justify the use of lagged variables in order to capture lagged effects that may be present in the data generating process. Accordingly, this note also demonstrates how simultaneity thwarts the identification of structural parameters associated with lagged effects. Finally, it highlights the role played by serial correlation in the endogenous explanatory variable, and demonstrates how this exacerbates problems of parameter recovery and hypothesis testing.

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

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 Xi,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 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).

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

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

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 the explanatory variable by f.

A similar problem arises if we regress 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') $\quad \Delta Y_t = \alpha + \beta X_{t-1} + \gamma Y_{t-1} + \ error\ term,$

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

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

2. The corresponding value of Xt is $X_t = \frac{(d+ae)}{(1-be)} + \frac{(f+ce)}{(1-be)} X_{t-1} + \frac{1}{(1-be)} \nu_t + \frac{e}{(1-be)} \varepsilon_t.$

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 opposite of *b* and *c*. The figure illustrates the case where $b=c=1$. Given these values, any value of $e>1$ produces $plim(\hat{\beta})<0$. We next consider three cases and identify the necessary 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* (β) tells him/her something about the direct effect of current *X* on current *Y*. From Equation (4), it follows that necessary conditions for $\hat{\beta}$ to consistently estimate $b\neq 0$ are (i) $c=0$, (ii) $e=0$ and (iii) $f=1$. 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 endogenous, so there is no simultaneity problem. The third condition is that *X*, and thus *Y*, are random walk processes. Only when all three conditions hold will OLS estimate $b\neq 0$ consistently, though the associated *t*-critical values will be invalid. Of course, if there is no simultaneity problem, there is no reason 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. The necessary conditions for $\hat{\beta}$ to consistently estimate the sum ($b+c$), $b\neq 0$, $c\neq 0$, are (i) $e=0$ and (ii) $f=1$. In other words, the strategy of substituting lagged *X* for current *X* will only produce consistent estimates of ($b+c$) when there is no simultaneity and *X* is a random walk; and then again conventional *t*-tests will be invalid. 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.

CASE THREE: *Estimation of c*. The third case represents the scenario where the researcher believes X affects Y with a lag. There are two sets of necessary conditions that will allow least squares regression to produce consistent estimates of $c \neq 0$. The first set of conditions are (i) $e=0$, and (ii) $f=0$. Under these conditions, there is no 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. The 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 attempting to avoid simultaneity by using lagged versus current values of a suspected endogenous variable. The only conditions where replacing X_t with X_{t-1} produces consistent estimates of the direct effects of X on Y is when there is no simultaneity.

III. SIMULATION RESULTS.

The preceding section demonstrates that lagging X does not enable one to escape simultaneity bias. In this section, we illustrate the risks associated with interpreting the coefficient on lagged X as a measure of the direct effect(s) of X on Y. The simulations also illustrate the important role played by serial correlation in the explanatory variable. This should be of interest given that many, if not most, economic time series are characterized by serial correlation.

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

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

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 are designed to represent the case where a researcher substitutes X_{t-1} for X_t for the sole reason of avoiding simultaneity bias.

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, $H_0: \beta = 0$, 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 zero as the sample sizes increase. The rejection rates for $H_0: \beta = 0$ are close to 0.05. This is all good as long as the researcher understands that he/she is estimating the direct effect of lagged X on Y, which in this case is zero. However, if the researcher believes that $\hat{\beta}$ measures either the direct effect of current X on Y (b), or the total direct effect of both X_t and X_{t-1} (b+c), he/she will incorrectly conclude that X has no effect on Y approximately 95% of the time.

SIMULATION 2 shows the effect that serial correlation has on the estimates. 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.

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 $\Delta Y_t = \alpha + \beta X_{t-1} + \gamma Y_{t-1} + error\ term$. 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.[3]

IV. CONCLUSION.

A common practice in applied econometrics work consists of replacing a suspected endogenous variable with its lagged values. This note demonstrates that lagging an endogenous variable 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. Indeed, 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. The only time a strategy of

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

replacing X_t with X_{t-1} enables consistent estimation of structural parameters is when there is no simultaneity to begin with.

There may be situations where lagging mitigates problems associated with simultaneity. However, these situations have not been identified in previous research. The implication of this research is that those who employ this approach should be explicit about the conditions under which it constitutes an effective estimation strategy.

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

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

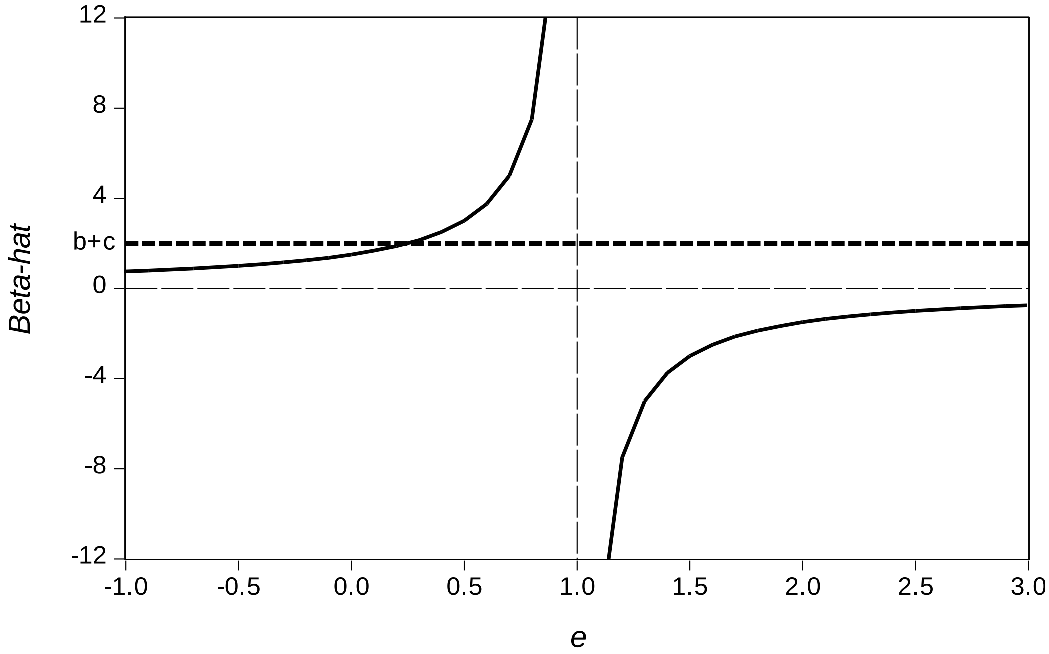
Nickell, S. J. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49: 1417–1426.

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 = 1
f = 0.5

NOTE: 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, 1, 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 of 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*.

TABLE 1
Simulation Results: Time Series Data

A) SIMULATION 1: ***Simultaneity but no Serial Correlation in the Endogenous Explanatory Variable (Dep. Var. in Level Form)***

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

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

	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 the Endogenous Explanatory Variable (Dep. Var. in Level Form)***

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

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

	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 the Endogenous Explanatory Variable (Dep. Var. in Difference Form)***

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

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

	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