ON THE PRACTICE OF LAGGING VARIABLES TO AVOID SIMULTANEITY

by

W. Robert Reed
Department of Economics and Finance
University of Canterbury
Christchurch, New Zealand

28 August 2013

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

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. The associated estimates are still distorted by simultaneity bias, and hypothesis testing is invalid. Further, I show that these problems are exacerbated when the suspected endogenous variable is characterised by serial correlation, as is likely often to be the case.

Keywords: Simultaneity, Endogeneity, Lagged variables

JEL classification: C1, C15, C2, C3
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 Coppelde (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 are heavily 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 the more transparent methods of lagging and differencing”1 (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).2

---

1 “Lagging” is the authors’ solution for reverse causation; “differencing” is designed to eliminate endogeneity from unobserved fixed effects.
2 Other examples are: “Innovation intensity is included as a lagged variable in order to mitigate simultaneity problems” (Aschoff 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 Coppelde, 2006, p. 476); “We lag the explanatory variables X_{lt-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”
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. Furthermore, the problem is exacerbated by the presence of serial correlation in the endogenous 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:

\[
Y_t = a + bX_t + cX_{t-1} + \varepsilon_t,
\]

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 simultaneously bias, the researcher estimates

\[
Y_t = \alpha + \beta X_{t-1} + \text{error term}.
\]

To determine the relationship between \( \beta \) 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

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

Then

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

It follows that OLS estimation of Equation (1’) produces a consistent estimate of the reduced form coefficient on lagged \( X \).

(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).

The corresponding value of \( X_t \) is

\[
X_t = \frac{(d+ae)}{(1-be)} \frac{(f+ce)}{(1-be)} X_{t-1} + \frac{1}{(1-be)} \nu_t + \frac{e}{(1-be)} \varepsilon_t .
\]

3
where simultaneity is represented by the parameter \( e \), and serial correlation in the explanatory variable by \( f \). In general, \( \text{plim}(\hat{\beta}) \neq c \), where \( c \) is the true effect of \( X_{t-1} \) on \( Y \).\(^4\)

A similar problem arises if we regress the change in \( Y \) on lagged \( X \). In this case,

\[
\Delta Y_t = \frac{(a + bd)}{(1-be)} + \frac{(c + bf)}{(1-be)} X_{t-1} - Y_{t-1} + \frac{b}{(1-be)} \nu_t + \frac{1}{(1-be)} \epsilon_t.
\]

OLS estimation of

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

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

\[
\text{plim}(\hat{\beta}) = \frac{(c + bf)}{(1-be)} \neq c.
\]

The preceding demonstrates that lagging \( X \) does not enable one to escape simultaneity bias. Furthermore, serial correlation in the explanatory variable further distorts parameter estimation and hypothesis testing. The next section uses Monte Carlo analysis to highlight these points.

### III. SIMULATION RESULTS

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

\[
\begin{align*}
Y_t &= 0.5 X_t + 0 X_{t-1} + \epsilon_t, \\
X_t &= 0.5 Y_t + f X_{t-1} + \nu_t, \\
\epsilon_t, \nu_t &\sim NID(0,1);
\end{align*}
\]

where \( f \) is alternatively set equal to 0 and 0.5. Note that \( Y_t \) and \( X_t \) are simultaneously determined, and that the true effect of \( X_{t-1} \) on \( Y_t \) is zero. When \( f = 0 \), there is no serial correlation in \( X \) (SIMULATION 1). However, when \( f = 0.5 \), \( X_t \) and \( X_{t-1} \) are correlated, with a mean sample correlation of approximately 0.63 (SIMULATION 2). For each case I use OLS to estimate the equation

\(^4\) It is sometimes stated that the use of lagged variables “reduces” or “mitigates” simultaneity bias. However, these statements have no meaning because the coefficients on \( X_t \) and \( X_{t-1} \) measure inherently different things.
6b) \( Y_t = \alpha + \beta X_{t-1} + \text{error term} \).

The simulated datasets vary in size from \( T=10 \) to \( T=1000 \). For each value of \( T \), 1000 data sets are generated, producing 1000 estimates of \( \beta \), the coefficient on \( X_{t-1} \) in Equation (6b). If \( X_t \) affects \( Y_t \), and \( X \) is serially correlated, then regression of \( Y_t \) on \( X_{t-1} \) will incorrectly estimate the effect of \( X_{t-1} \) on \( Y_t \). The table reports the mean estimate of \( \beta \) for each set of replications, along with the rate at which the null hypothesis, \( H_0: \beta = 0 \), is rejected. Rejection rates larger than 0.05 would cause the researcher to too-frequently conclude that \( X_{t-1} \) has a direct impact on \( Y_t \).

As noted above, the only difference between SIMULATION 1 and SIMULATION 2 is serial correlation in \( X \). In SIMULATION 1, the mean estimated value of \( \beta \) suffers from finite sample bias, but converges to zero as the sample sizes increase in size. The rejection rates for \( H_0: \beta = 0 \) are all close to 0.05. In contrast, in the presence of serial correlation (SIMULATION 2), the estimates of \( \beta \) converge to their reduced form value of \( \frac{c+bf}{1-be} = 0.33 \). This causes the researcher to incorrectly conclude that \( X_{t-1} \) has a direct impact on \( Y_t \) in more than 5 percent of the estimated equations. In fact, the rejection rate of the null converges to 1.000 as the sample size increases. Note that there are two sources contributing to this error: (i) serial correlation in \( X \), and (ii) the researcher’s misguided attempt to attach a structural interpretation to a reduced-form expression.

SIMULATION 3 repeats the analysis of SIMULATION 2 except that the dependent variable in the DGP is \( \Delta Y_t \) rather than \( Y_t \), and the estimated equation is now \( \Delta Y_t = \alpha + \beta X_{t-1} + \gamma Y_{t-1} + \text{error term} \). 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 \( \beta \) converge to their reduced form value of \( \frac{c+bf}{1-be} = 0.33 \), and the rejection rate of the null converges to 1.000 as the sample size increases.
The next set of simulations investigates the same type of relationships, except that I now work with panel data. In particular, I set the number of cross-sectional units \(N\) equal to 50, and let the sample sizes range from \(T=10\) to \(T=100\). The DGP is now:

7a) \[ Y_{it} = 0.5 X_{it} + 0 X_{i,t-1} + \lambda_i + \varepsilon_{it} \]
\[ X_{it} = 0.5 Y_{it} + f X_{i,t-1} + \mu_i + \nu_{it} \]
where \(f\), again, is alternatively set equal to 0 and 0.5. The estimated equation is now

7b) \[ Y_{it} = \alpha + \beta X_{i,t-1} + \text{fixed effects} + \text{error term}, \]
which is estimated using OLS with fixed effects. The corresponding results are reported in TABLE 2.

SIMULATION 4 is the case with no serial correlation \((f = 0)\). It is noteworthy that the Mean \(\hat{\beta}\) values are similar to the time series results of SIMULATION 1 -- another “no serial correlation” case -- for similar \(T\) values. However, the standard errors are smaller given the overall larger sample sizes \((NT)\), so that the rejection rates are substantially greater than 0.05 at small \(T\). As \(T\) gets larger, the rejection rates converge to their “true” value of 0.05.

When serial correlation is present (SIMULATION 5), the values for Mean \(\hat{\beta}\) are again similar to their non-panel matches for like \(T\) (cf. SIMULATION 2), but the rejection rates converge to 1.000 much faster. By \(T=20\), the null hypothesis of \(H_0: \beta = 0\), is rejected 100 percent of the time.

SIMULATIONS 6 and 7 repeat the panel data analysis of SIMULATIONS 4 and 5, except the dependent variable is now \(\Delta Y_t\) rather than \(Y_t\). The corresponding estimated equation is \(\Delta Y_{it} = \alpha + \beta X_{i,t-1} + \gamma Y_{i,t-1} + \text{fixed effects} + \text{error term}\). The combination of a lagged dependent variable and fixed effects raises concerns of “dynamic panel bias” (Nickell, 1981), which is problematic for large \(N\), small \(T\) data. Given \(N = 50\), I investigate the consequences of serial correlation for the cases \(T = 10\) and \(T = 20\). I estimate the model
using difference GMM (Arellano and Bond, 1991). Note that difference GMM transforms \( \Delta Y_t = \alpha + \beta X_{i,t-1} + \gamma Y_{i,t-1} + \text{fixed effects} \) to \( \Delta \Delta Y_t = \beta \Delta X_{i,t-1} + \gamma \Delta Y_{i,t-1} \), where \( \Delta X_{i,t-1} \) and \( \Delta Y_{i,t-1} \) are instrumented by past levels of \( X \) and \( Y \) (we assume the researcher would consider both to be predetermined, but not endogenous). This is an important difference with previous estimation, because \( \Delta X_{i,t} \) is effectively replaced by its instrumented predicted value.

TABLE 3 reports the corresponding Mean \( \hat{\beta} \) and rejection rate results for the “no serial correlation” and “serial correlation” cases, respectively. While there are some interesting differences, the main conclusions from the preceding analysis continues to hold.

**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. The associated estimates are still distorted by simultaneity bias, and hypothesis testing is invalid. Further, I show that these problems are exacerbated when the suspected endogenous variable is characterised by serial correlation, as is likely often to be the case.

---

5 Estimates were obtained using the xtabond2 procedure using the following commands:
\[
\text{xtabond2 D.y L.x L.y, gmm(L.(x y)) nolevel twostep robust small.}
\]

6 While not reported, we also used the system GMM procedure. For the serial correlation case, the coefficient estimates were closer to their reduced-form, \( \text{plim} \) values; and the rejection rates were somewhat higher.
REFERENCES


### 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) \( Y_t = 0.5 \ X_t + 0 \ X_{t-1} + \varepsilon_t \)
2) \( X_t = 0.5 \ Y_t + 0 \ X_{t-1} + \nu_t \), \( \varepsilon_t, \nu_t \sim \text{NID}(0,1) \)

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

<table>
<thead>
<tr>
<th>T=10</th>
<th>T=20</th>
<th>T=50</th>
<th>T=100</th>
<th>T=1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \hat{\beta} )</td>
<td>-0.0728</td>
<td>-0.0309</td>
<td>-0.0093</td>
<td>-0.0039</td>
</tr>
<tr>
<td>Rejection Rate for H0: ( \beta = 0 )</td>
<td>0.039</td>
<td>0.047</td>
<td>0.051</td>
<td>0.051</td>
</tr>
</tbody>
</table>

**B) SIMULATION 2: Simultaneity with Serial Correlation in the Endogenous Explanatory Variable (Dep. Var. in Level Form)**

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

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

<table>
<thead>
<tr>
<th>T=10</th>
<th>T=20</th>
<th>T=50</th>
<th>T=100</th>
<th>T=1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \hat{\beta} )</td>
<td>0.1198</td>
<td>0.2202</td>
<td>0.2925</td>
<td>0.3123</td>
</tr>
<tr>
<td>Rejection Rate for H0: ( \beta = 0 )</td>
<td>0.071</td>
<td>0.234</td>
<td>0.701</td>
<td>0.951</td>
</tr>
</tbody>
</table>

**C) SIMULATION 3: Simultaneity with Serial Correlation in the Endogenous Explanatory Variable (Dep. Var. in Difference Form)**

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

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

<table>
<thead>
<tr>
<th>T=10</th>
<th>T=20</th>
<th>T=50</th>
<th>T=100</th>
<th>T=1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \hat{\beta} )</td>
<td>0.1634</td>
<td>0.2478</td>
<td>0.3087</td>
<td>0.3185</td>
</tr>
<tr>
<td>Rejection Rate for H0: ( \beta = 0 )</td>
<td>0.067</td>
<td>0.113</td>
<td>0.348</td>
<td>0.671</td>
</tr>
</tbody>
</table>
TABLE 2
Simulation Results: Panel Data (Dependent Variable in Level Form)

A) SIMULATION 4: Simultaneity but no Serial Correlation in the Endogenous Explanatory Variable

DGP:

1) $Y_{it} = 0.5 X_{it} + 0 X_{i,t-1} + \lambda_i + \epsilon_{it}$
2) $X_{it} = 0.5 Y_{it} + 0 X_{i,t-1} + \mu_i + \nu_{it}$

$\lambda_i = \mu_i = 0$, $i = 1, 2, \ldots, 50$; $\epsilon_{it}, \nu_{it} \sim NID(0,1)$

Estimated Equation (OLS): $Y_{it} = \alpha + \beta X_{i,t-1} + \text{fixed effects} + \text{error term}$

<table>
<thead>
<tr>
<th></th>
<th>$T=10$</th>
<th>$T=20$</th>
<th>$T=50$</th>
<th>$T=100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $\hat{\beta}$</td>
<td>-0.0869</td>
<td>-0.0410</td>
<td>-0.0161</td>
<td>-0.0080</td>
</tr>
<tr>
<td>Rejection Rate for $H_0$: $\beta = 0$</td>
<td>0.404</td>
<td>0.231</td>
<td>0.127</td>
<td>0.091</td>
</tr>
</tbody>
</table>

B) SIMULATION 5: Simultaneity with Serial Correlation in the Endogenous Explanatory Variable

DGP:

1) $Y_{it} = 0.5 X_{it} + 0 X_{i,t-1} + \lambda_i + \epsilon_{it}$
2) $X_{it} = 0.5 Y_{it} + 0.5 X_{i,t-1} + \mu_i + \nu_{it}$

$\lambda_i = \mu_i = 0$, $i = 1, 2, \ldots, 50$; $\epsilon_{it}, \nu_{it} \sim NID(0,1)$

Estimated Equation (OLS): $Y_{it} = \alpha + \beta X_{i,t-1} + \text{fixed effects} + \text{error term}$

<table>
<thead>
<tr>
<th></th>
<th>$T=10$</th>
<th>$T=20$</th>
<th>$T=50$</th>
<th>$T=100$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean $\hat{\beta}$</td>
<td>0.1626</td>
<td>0.2564</td>
<td>0.3043</td>
<td>0.3191</td>
</tr>
<tr>
<td>Rejection Rate for $H_0$: $\beta = 0$</td>
<td>0.934</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
### TABLE 3
Simulation Results: Panel Data (Dependent Variable in Difference Form)

#### A) SIMULATION 6: Simultaneity but no Serial Correlation in the Endogenous Explanatory Variable

DGP:
1) \[ \Delta Y_{it} = 0.5 X_{it} + 0 X_{i,t-1} - Y_{i,t-1} + \lambda_i + \epsilon_{it} \]
2) \[ X_{it} = 0.5 Y_{it} + 0 X_{i,t-1} + \mu_i + \nu_{it} \]
\[ \lambda_i = \mu_i = 0 , i = 1,2,...,50; \epsilon_{it},\nu_{it} \sim NID(0,1) \]

Estimated Equation (Difference GMM):
\[ \Delta Y_{it} = \alpha + \beta X_{i,t-1} + \gamma Y_{i,t-1} + \text{fixed effects} + \text{error term} \]

<table>
<thead>
<tr>
<th>T=10</th>
<th>T=20</th>
<th>T=50</th>
<th>T=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \hat{\beta} )</td>
<td>0.0013</td>
<td>0.0256</td>
<td>----</td>
</tr>
<tr>
<td>Rejection Rate for ( H_0: \beta = 0 )</td>
<td>0.044</td>
<td>0.001</td>
<td>----</td>
</tr>
</tbody>
</table>

#### B) SIMULATION 7: Simultaneity with Serial Correlation in the Endogenous Explanatory Variable

DGP:
1) \[ \Delta Y_{it} = 0.5 X_{it} + 0 X_{i,t-1} - Y_{i,t-1} + \lambda_i + \epsilon_{it} \]
2) \[ X_{it} = 0.5 Y_{it} + 0.5 X_{i,t-1} + \mu_i + \nu_{it} \]
\[ \lambda_i = \mu_i = 0 , i = 1,2,...,50; \epsilon_{it},\nu_{it} \sim NID(0,1) \]

Estimated Equation (Difference GMM):
\[ \Delta Y_{it} = \alpha + \beta X_{i,t-1} + \gamma Y_{i,t-1} + \text{fixed effects} + \text{error term} \]

<table>
<thead>
<tr>
<th>T=10</th>
<th>T=20</th>
<th>T=50</th>
<th>T=100</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean ( \hat{\beta} )</td>
<td>0.2792</td>
<td>0.3008</td>
<td>----</td>
</tr>
<tr>
<td>Rejection Rate for ( H_0: \beta = 0 )</td>
<td>0.745</td>
<td>0.615</td>
<td>----</td>
</tr>
</tbody>
</table>