**DEPARTMENT OF ECONOMICS AND FINANCE
COLLEGE OF BUSINESS AND ECONOMICS
UNIVERSITY OF CANTERBURY
CHRISTCHURCH, NEW ZEALAND**

**On Estimating Long-Run Effects
in Models with Lagged Dependent Variables**

**W. Robert Reed
Min Zhu**

WORKING PAPER

No. 18/2015

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

WORKING PAPER No. 18/2015

On Estimating Long-Run Effects in Models with Lagged Dependent Variables

W. Robert Reed^{1*}
Min Zhu²

November 30, 2015

Abstract: This note points out the hazards of estimating long-run effects from models with lagged dependent variables. We use Monte Carlo experiments to demonstrate that this practice often fails to produce reliable estimates. Biases can be substantial, sample ranges very wide, and hypothesis tests can be rendered useless in realistic data environments. There are three reasons for this poor performance. First, OLS estimates of the coefficient of a lagged dependent variable are downwardly biased in finite samples. Second, small biases in the estimate of the lagged, dependent variable coefficient are magnified in the calculation of long-run effects. And third, and perhaps most importantly, the statistical distribution associated with estimates of the *LRP* is complicated, heavy-tailed, and difficult to use for hypothesis testing. While alternative procedures such as jackknifing and indirect inference address the first issue, associated estimates of long-run effects remain unreliable.

Keywords: Hurwicz bias, Auto-Regressive Distributed-Lag models, ARDL, Dynamic Panel Data models, DPD, Anderson-Hsaio, Arellano-Bond, Difference GMM, System GMM, indirect inference, jackknifing, long-run impact, long-run propensity

JEL Classifications: C22, C23

Acknowledgments: This paper acknowledges helpful comments and discussions with David Frazier, David Giles, Andrew Harvey, Adrian Pagan, Yongcheol Shin, Richard J. Smith, Jeffrey Wooldridge, Huanjun Zhu; conference participants at the New Zealand Econometric Study Group in Oz meeting (February 2015), the New Zealand Association of Economists meeting (July 2015), the Advances in Empirical and Theoretical Econometrics Workshop (November 2015); and seminar participants at the University of Otago and Waikato University.

¹ Department of Economics and Finance, University of Canterbury, Christchurch, NEW ZEALAND

² School of Economics and Finance, Queensland University of Technology, AUSTRALIA

I. INTRODUCTION

This note points out the hazards of estimating long-run effects from models with lagged dependent variables. We illustrate the problem in the context of the following ARDL(p, q) model where $p=1$ and $q=0$:

$$(1) \quad y_t = \beta_0 + \beta_x x_t + \beta_y y_{t-1} + u_t, u_t \sim NID(0, \sigma^2), t = 1, 2, \dots, T.$$

We assume $|\beta_y| < 1$. It is common within this framework to estimate the long-run effect of x on y , also known as the long-run propensity (*LRP*) of x , by

$$(2) \quad LRP = \hat{\beta}_x / (1 - \hat{\beta}_y),$$

where $\hat{\beta}_x$ and $\hat{\beta}_y$ are the OLS estimates from Equation (1). Generalization to higher orders of p and q is straightforward.

Examples of this procedure for calculating long-run effects are easy to find. It appears in a large number of econometrics textbooks, including Judge et al. (1988, page 737), Johnston & Dinardo (1997, page 245), Hill et al. (2001, page 328), Dougherty (2002, page 321), Kennedy (2008, page 322), Gujarati and Porter (2010, page 378), Greene (2012, page 423), and Wooldridge (2012, page 635). And it shows up frequently in economics journals. TABLE 1 lists a sampling of articles that calculate long-run effects in this manner. It includes a wide variety of journals, such as *Review of Economics and Statistics*; *Journal of Money, Credit and Banking*; *Journal of Development Economics*; and *Economics Letters*; among others. Long-run effects are calculated both from single-equation ARDL models (OLS), and dynamic panel data (DPD) models. Many of the time periods are relatively short.

Unfortunately, this procedure has serious statistical pitfalls. It is well known that OLS estimation of autoregressive (AR) models -- while consistent -- is biased in finite samples (Hurwicz, 1950; Phillips, 1977). It follows that estimates of the *LRP* will likewise be biased. Even worse, the *LRP* is a nonlinear function of the AR coefficient(s), and a small bias in the denominator can be greatly magnified, especially when the AR coefficients are

close to the unit circle. Marriott and Pope (1954) illustrate this in the case when $p = 1$. The first-order bias for OLS estimation of β_y is $-(1 + 3\beta_y)/T$. When $\beta_y = 0.95$ and $T = 100$, this implies a bias of -0.0385, or 4% of the true value. However, the associated bias in the *LRP* is 43.5%, since the corresponding estimate of *LRP* is 11.3, compared to its true value of 20. Things can get even worse when higher order AR terms are included in the equation. As pointed out by Patterson (2000), small biases in the estimates of the individual AR coefficients cumulate, often in a reinforcing rather than offsetting manner.

And there is yet another issue. The *LRP* of Equation (2) is the ratio of two normally distributed random variables. As is well known, when both random variables are independent and have zero means, their ratio is distributed Cauchy. However, when the two variables have non-zero means, Hinkley (1969) has shown that the associated distribution is quite complicated. The shape of its density can be unimodal, bimodal, symmetric, or asymmetric depending on the value of the coefficient of the denominator variable. Due to this fact, it is difficult to work with the *LRP* and develop appropriate statistical tests.

At some level, most of the above is already known. What is not known – as evidenced by the widespread use of this procedure – is the practical importance of these problems for estimates of long-run effects in realistic data environments. That is the contribution of this study. We conduct a variety of Monte Carlo simulation experiments and demonstrate that these problems can be severe. We then extend our analysis and apply recent time series procedures (jackknifing and indirect inference) to see if they alleviate the problem. We show that they do not.

The study proceeds as follows. Section 2 sets the context by illustrating the finite sample bias associated with OLS estimation of a simple AR(1) model. Section 3 expands the analysis to include the estimation of the *LRP* in the framework of an ARDL model. Section 4 further extends the analysis to a panel data framework where three different DPD estimators

are used to estimate the *LRP*. Section 5 revisits the estimation of the ARDL and DPD models using jackknifing and indirect inference. Section 6 concludes by drawing implications of the preceding analysis for the estimation of long-run effects.

2. FINITE-SAMPLE BIAS IN OLS ESTIMATION OF AN AR MODEL

We set the context by using Monte Carlo experiments to study the model:

$$(3) \quad y_t = 1 + \beta_y y_{t-1} + u_t, t = 1, 2, \dots, T, u_t \sim NID(0, 1).$$

The error terms u_t are generated independently of the lagged dependent variable values, y_{t-1} . T takes values between 15 and 1000 ($T = 15, 30, 50, 100$, and 1000), and β_y takes values between 0.010 and 0.99 ($\beta_y = 0.10, 0.50, 0.90, 0.95$, and 0.99). For each simulated sample, OLS is used to estimate the value of β_y . 10,000 replications were run for each of the (T, β_y) combinations.

The results of the Monte Carlo experiments are reported in TABLE 2.¹ The top panel of the table reports the average estimated value of $\hat{\beta}_y$ across the 10,000 replications. The bottom panel reports Type I error rates, where the values are the rejection rates of the null hypothesis, $H_0: \beta_y = \text{its true value}$. For both panels, the columns represent the different sample sizes, and the rows represent the different true values of β_y .

For example, when $T=15$ and $\beta_y = 0.90$, the average estimated value of $\beta_y = 0.640$, substantially below its true value. When $T=1000$, the average estimated value of β_y increases to 0.896, illustrating the consistency of the OLS estimator. Likewise, when $T=15$ and $\beta_y = 0.90$, the rejection rate of the null hypothesis $H_0: \beta_y = 0.90$, given a significance level of 5%, is 0.131. This declines as the sample size gets larger, falling to 0.055 when $T=1000$.

Note that most of the ARDL/OLS studies in TABLE 1 have fewer than 50 time observations. TABLE 2 suggests that this can be a problem. For example, when $T=50$ and

¹ Tables 2 through 5 were produced using Stata, Version 13. Tables 6 and 7 were produced using Matlab, and Table 8 used R. All programs are available from the authors upon request.

$\beta_y = 0.90$, the average estimated value of β_y is 0.819, with a nominal bias of -0.081, or -9.0 percent.² Similarly, the rejection rate of the null hypothesis $H_0: \beta_y = 0.90$ is 0.103 (significance level is 5%). This value rises substantially as β_y gets closer to one.

3. ESTIMATION OF LONG-RUN EFFECTS IN AN ARDL MODEL

The next set of Monte Carlo experiments add an explanatory variable to the data generating process (DGP) of Equation (3):

$$(4) \quad y_t = 1 + \beta_x x_t + \beta_y y_{t-1} + u_t, \quad t = 1, 2, \dots, T, \quad u_t \sim NID(0, 1),$$

where $\beta_x = 1$ and β_y takes values 0.10, 0.50, 0.90, 0.95, and 0.99 (same as above). The corresponding, true *LRP* values are 1.11, 2, 10, 20, and 100. The estimated *LRP* values are calculated by $\hat{\beta}_x / (1 - \hat{\beta}_y)$.

TABLE 3 reports the results of these Monte Carlo experiments. The table is organized similarly to TABLE 2 with some minor differences. We now report three panels. The top panel reports the *median* (not average) value of the estimated *LRP* values across the 10,000 replications.³ The middle panel reports the 90 percent empirical sample range calculated by taking the 5th and 95th percentile values of the 10,000 estimated *LRP* values. The bottom panel reports Type I error rates associated with the null hypothesis, $H_0: LRP = \text{its true value}$. While the range of the true β_y values is the same as above, we only report simulation results for sample sizes of $T = 10, 50$, and 1000 .

There is little nominal distortion associated with small values of β_y when $T = 50$. For example, when $\beta_y = 0.10$, the median estimated value of *LRP* is 1.11, the same as the true value. The 90 percent empirical sample range extends from 0.08 to 1.5, encompassing the true value. And the associated Type I error rate is 0.064.

² Note that the percentage bias is often larger for smaller values of β_y .

³ I report the median rather than the mean, because estimates of β_y very close to 1 can cause the *LRP* to be extremely large, making the sample mean uninformative.

However, substantial distortions arise for larger values of β_y . For example, when the sample size remains at 50 and β_y increases to 0.90, so that the true value of $LRP = 10$, the median estimated LRP value is 7.5, a bias of negative 25 percent. The 90 percent empirical sample range, while encompassing the true value, is wider, extending from 3.4 to 19.4. And the associated Type I error rate is 0.267. All of these values are substantially worse for larger values of β_y and/or smaller values of T .

4. ESTIMATION OF LONG-RUN EFFECTS IN A DYNAMIC PANEL DATA MODEL

In this section, we extend the DGP of the previous section by incorporating the model in a panel data setting. The corresponding DGP is:

$$(5) \quad y_{it} = 1 + \beta_x x_{it} + \beta_y y_{i,t-1} + \varepsilon_{it}, \quad i = 1, 2, \dots, N, t = 1, 2, \dots, T,$$

where $\beta_x = 1$ and β_y again takes values 0.10, 0.50, 0.90, 0.95, and 0.99. ε_{it} is comprised of a fixed effect and a classical error term, $\varepsilon_{it} = a_i + u_{it}$, where each of the components of ε_{it} are independently and identically standard normally distributed, $a_i \sim N(0,1)$, $u_{it} \sim NID(0,1)$. The fixed effects are uncorrelated with x_{it} , so there is no endogeneity associated with the fixed effects. We begin with the panel data having dimensions $N=50, T=10$.

Given the values of β_y , the true LRP values are, again, 1.11, 2, 10, 20, and 100, respectively. We estimate the model using three dynamic panel data (DPD) estimators: Anderson-Hsaio (AH), difference GMM (DGMM), and system GMM (SGMM). The results are reported in TABLE 4 and again consist of three panels: (i) the median estimated LRP values, (ii) the 90% empirical sample ranges of LRP values, and (iii) the Type I error rates. The columns now report the results for each of the three DPD estimators. Note that sample sizes are the same for all the experiments reported in TABLE 4 ($N=50, T=10$).

Substantial biases in the LRP estimates are again evident for large values of β_y . For example, when $\beta_y = 0.90$, the median LRP estimates for the AH, DGMM, and SGMM

estimators are 0.6, 5.1, and 25.6 (compared to a true value of 10). The 90% empirical sample ranges are very wide, and in none of the cases do they include the true value. The Type 1 error rates are 0.709, 0.821, and 0.001, rendering them worthless for hypothesis testing. Larger values of β_y result in even worse performance.

TABLE 5 reports a similar analysis for panel data having dimensions $N=140$, $T=5$. We chose these latter dimensions because they are the approximate dimensions of the “abdata” dataset bundled with Stata that is used to illustrate the functionality of the *xtabond* (difference GMM) and the *xtdpdsys* (system GMM) commands in Stata.⁴ The results are very similar to those from TABLE 4. Once again, substantial biases and size distortions are apparent for large values of β_y .

5. JACKKNIFING AND INDIRECT INFERENCE

Jackknifing (JK) and indirect inference (II) are two methods that have been suggested for correcting the bias associated with estimating AR/ARDL models in finite samples (for example, see Dhaene and Jochmans, 2015; Gouriéroux et al., 2010). In this section, we apply these methods in order to obtain less biased estimates of β_y . We then investigate whether improves the corresponding *LRP* estimates.

The jackknife estimator removes bias by comparing the maximum-likelihood estimate from the full sample with estimates calculated from subsamples. Our analysis partitions the full sample into two subsamples. We briefly describe our procedure. Let $\hat{\beta}$ be the maximum-likelihood based estimator of β . Partition the observations along the time dimension. Given $S = \{1, \dots, T\}$, create two subsamples, $S_1 = \{1, \dots, \frac{T}{2}\}$ and $S_2 = \{\frac{T}{2} +$

⁴ The “abdata” data set used for the *xtabond* procedure has 140 groups, with a minimum of 4 observations per group, and a maximum of 6. The corresponding minimum and maximum values corresponding to the data used for the *xtdpdsys* procedure are 5 and 7, with the number of groups again equaling 140.

$1, \dots, T\}$. Let $\bar{\beta}_{1/2} = 0.5(\hat{\beta}_{S_1} + \hat{\beta}_{S_2})$. Compared to the original estimator, $\hat{\beta}$, the jackknifed estimator $\bar{\beta}_{1/2}$ will have smaller bias.

Indirect inference is a simulation-based method for bias reduction. The following is a brief description of the II procedure we employ. Suppose the true data generating process takes the form $y_t = F(y_{t-1}, x_t | \theta)$. In our case, θ is a vector (β, σ) , where β is the regression coefficients and σ is the standard error of the regression error term. Suppose that $\hat{\theta} = (\hat{\beta}, \hat{\sigma})$ is an estimator of θ based on observed data. The II procedure is carried out in three steps. Step 1 generates a sequence of random errors of length T from the F distribution conditional on the estimated parameter $\hat{\theta}$. Step 2 chooses values for the parameter vector β using $\{x_t\}_{t=1}^T$ and the simulated random errors in Step 1 to generate the sequence $\{\tilde{y}_t(\beta)\}_{t=1}^T$. The $\{\tilde{y}_t(\beta)\}_{t=1}^T$ and $\{x_t\}_{t=1}^T$ values are then used to calculate an estimate of β , $\delta(\beta)$. Step 3 repeats Steps 1 and 2 M times (M simulations) for a given set of β values, producing the estimates $\{\delta_m(\beta)\}_{m=1}^M$. The bias-reduced estimator is obtained by $\hat{\beta}_{II} = \text{argmin}_{\beta} (\hat{\beta} - \frac{1}{M} \sum_{m=1}^M \delta_m(\beta))$. The central idea of indirect inference is to choose β so that $\delta(\beta)$ and $\hat{\beta}$ are as close as possible.

We first demonstrate that jackknifing and indirect inference are effective at reducing the bias associated with estimating β_y in Equation (4). The top panel of TABLE 6 reports the OLS estimates of β_y . The next two panels report the JK and II estimates. As before, the Monte Carlo experiments have β_y take values 0.10, 0.50, 0.90, 0.95, and 0.99. The sample sizes range from $T=10$ to $T=1000$.

As was foreshadowed by TABLE 2, OLS estimates of β_y in the ARDL model are downwardly biased in finite samples. For example, when $T=10$ and $\beta_y = 0.90$, the average OLS estimate of β_y is 0.69. In contrast, the corresponding, average JK and II estimates are 0.85 and 0.86. That these latter two estimates are very close to the true value of β_y even for

$T=10$ is testimony to the effectiveness of these procedures to reduce finite sample bias. The JK and II estimates of β_y weakly dominate the OLS estimates across the full range of β_y and T values. However, as we show next, these performance improvements do not translate into better estimates of LRP .

The LRP estimates for the JK and II estimators are reported in TABLE 7. These can be compared to the OLS estimates of TABLE 3. TABLE 7 shows improvement in some individual cases. For example, when $\beta_y = 0.90$ and $T=50$ and the true LRP value is 10, the JK and II estimates are 8.2 and 9.9, compared to an OLS estimate of 7.5. However, there are other instances when the JK and II estimates are worse than the OLS estimates. For example, when $\beta_y = 0.99$ and $T=50$ and the true LRP value is 100, the JK and II estimates are 6.8 and 8.4, compared to an OLS estimate of 16.0.

Furthermore, even when the JK and II estimates of LRP are less biased than OLS, they often have greater dispersion. For example, in the previous case of $\beta_y = 0.90$ and $T=50$, the 90 percent empirical sample ranges of the JK and II estimates are (-35.1, 52.0) and (-36.5, 55.2). These compare to (3.4, 19.4) for OLS. Similarly, when $\beta_y = 0.99$ and $T=1000$, the JK and II sample ranges are (49.2, 330.1) and (56.8, 249.4), compared to an OLS sample range of (51.2, 152.9). Overall, we conclude that the bias reductions provided by JK and II in the estimation of β_y do not translate to substantial improvements in the bias and range of the LRP estimates. A similar conclusion applies to the hypothesis test results, as evidenced by a comparison of the bottom panels of TABLES 7 and 3.

The reason for the poor performance of the JK and II estimators has to do with the unusual nature of the statistical distribution of $\hat{\beta}_x/(1 - \hat{\beta}_y)$. As noted above, this ratio is a nonlinear function of $\hat{\beta}_y$. Further, the distribution is heavy-tailed, non-standard, and complicated. The shape of its density can be unimodal, bimodal, symmetric, or asymmetric.

As a result, there is no reason to expect that just because the estimate of β_y is less biased that the estimate of $\hat{\beta}_x/(1 - \hat{\beta}_y)$ will be as well.

TABLE 8 reports *LRP* estimates associated with estimating the DPD model ($N=50$, $T=10$) via jackknifing and indirect inference. A comparison of TABLE 8 to TABLE 4 reinforces the previous conclusion: while jackknifing and indirect inference reduce the finite sample bias associated estimating β_y , this does not translate into improved *LRP* estimates.

6. CONCLUSION

A common practice in economic research is to calculate long-run impacts based on the estimated coefficients from a model involving one or more lagged dependent variables. For example, given the model $y_t = \beta_0 + \beta_x x_t + \beta_y y_{t-1} + u_t$, the long-run effect of x on y , also known as the long-run propensity (*LRP*) of x , is estimated by

$$(2) \quad LRP = \hat{\beta}_x / (1 - \hat{\beta}_y),$$

This note points out that this practice is hazardous. It often fails to produce reliable estimates. Biases can be substantial, sample ranges very wide, and hypothesis tests can be rendered useless in realistic data environments.

The reason for this poor performance is threefold. First, OLS estimates of the coefficient of a lagged dependent variable are downwardly biased in finite samples (Hurwicz, 1950; Phillips, 1977). Second, small biases in the estimate of β_y will be substantially magnified in estimates of *LRP* when the value of β_y is close to one. And perhaps most importantly, the statistical distribution associated with estimates of *LRP* is heavy tailed and complicated. While some of these problems have previously been identified in the literature, this procedure continues to be presented in econometrics textbooks, and appear in top economics journals. Further, all of the problems identified above occur in a relatively

uncomplicated data environment. For example, the presence of serial correlation in the error term is likely to compound the problems identified here.⁵

There are a number of areas that could be explored by future research. An alternative to using a lagged dependent variable is to replace the dependent variable with lagged values of the explanatory variable(s). While this has its own set of shortcomings, it may be that it is possible to identify situations, and tests, to indicate when this strategy is likely to produce superior estimates. The effect of serial correlation in the DGP on *LRP* estimates is also an area worth investigating. Until such time that a solution to this problem can be found, we recommend that researchers acknowledge that reported estimates of long-run estimates are likely to be biased and unreliable.

⁵ The following papers provide useful discussions of the problems associated with serially-correlated error terms in dynamic models: Griliches (1961), Keele and Kelly (2006), Beck and Katz (2011) and Wilkins (2015).

REFERENCES

- Baltagi, B. H., Demetriades, P. O., and Law, S. H. (2009). Financial development and openness: Evidence from panel data. *Journal of Development Economics*, Vol. 89(2), pp. 285-296.
- Beck, N. and Katz, J.N. 2011. "Modelling Dynamics in Time-Series-Cross-Sectional Political Economy Data." *Annual Review of Political Science* Vol. 14, pp. 331–52.
- Bentzen, J. and Engsted, T. (2001). A revival of the autoregressive distributed lag model in estimating energy demand relationships. *Energy*, Vol. 26, pp. 45-55.
- Carstensen, K. and Toubal, F. Foreign direct investment in Central and Eastern European countries: a dynamic panel analysis. *Journal of Comparative Economics*, Vol. 32, pp. 3-22.
- Dergiades, T., and Tsoulfidis, L. (2008). Estimating residential demand for electricity in the United States, 1965–2006. *Energy Economics*, Vol. 30(5), pp. 2722-2730.
- De Vita, G. (2014). The long-run impact of exchange rate regimes on international tourism flows. *Tourism Management*, Vol. 45, pp. 226-233.
- Diaz-Frances, E. and Rubio, F.J. 2013. "On the Existence of a Normal approximation to the distribution of the Ratio of Two Independent Normal Random Variables." *Statistical Papers*, Vol. 54(2), pp. 309-323.
- Dhaene, G. and Jochmans, K. 2015. "Split-panel jackknife estimation of fixed-effect models." *Review of Economic Studies*, forthcoming.
- Dougherty, C. (2002). *Introduction to econometrics* (2nd ed.). Oxford: Oxford University Press.
- Greene, W. H. (2012). *Econometric analysis* (7th ed.). Upper Saddle River, N.J: Prentice Hall.
- Gourieroux, C., Phillips, P.C.B. and Yu, J. (2010). "Indirect inference for dynamic panel models" *Journal of Econometrics*, Vol 157, pp. 68–77 .
- Griliches, Zvi. 1961. "A Note of Serial Correlation Bias in Estimates of Distributed Lags." *Econometrica* Vol. 29, pp. 65–73.
- Gujarati, D. N., and Porter, D. C. (2010). *Essentials of Econometrics* (4th ed.). New York: McGraw-Hill/Irwin.
- Heid, B., Langer, J., & Larch, M. (2012). Income and democracy: Evidence from system GMM estimates. *Economics Letters*, Vol. 116(2), pp. 166-169.

- Hill, R. C., Griffiths, W. E., and Judge, G. G. (2001). *Undergraduate econometrics* (2nd ed.). New York: Wiley.
- Hinkley, D.V. 1969. "On the Ratio of Two Correlated Normal Random Variables." *Biometrika*, Vol. 56 No. 3, pp. 635-639.
- Hoque, M. M., and Yusop, Z. (2010). Impacts of trade liberalisation on aggregate import in Bangladesh: An ARDL Bounds test approach. *Journal of Asian Economics*, Vol. 21(1), pp. 37-52.
- Horioka, C. Y., and Wan, J. (2007). The determinants of household saving in China: a dynamic panel analysis of provincial data. *Journal of Money, Credit and Banking*, Vol. 39, pp. 2077-2096.
- Hurwicz, L. 1950. "Least Squares Bias in Time Series," in *Statistical Inference in Dynamic Economic Models*, ed. by T. C. Koopmans. New York: Wiley.
- Islam, F., Shahbaz, M., Ahmed, A. U., & Alam, M. M. (2013). Financial development and energy consumption nexus in Malaysia: A multivariate time series analysis. *Economic Modelling*, Vol. 30, pp. 435-441.
- Johnston, J. and Dinardo, J. (1997). *Econometric methods* (4th ed.). New York: McGraw-Hill.
- Judge, G. G., Griffiths, W.E., Hill, R.C., Lütkepohl, H., and Lee, T-C. (1988). *Introduction to the theory and practice of econometrics* (2nd ed.). New York: Wiley.
- Keele, L. and Kelly, N.J. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* Vol. 14: pp. 186-205.
- Kennedy, P. (2008). *A Guide To Econometrics* (6th ed.). New York: Wiley-Blackwell.
- Li, T., Lai, J.T., Wang, Y., and Zhao, D. (2015). Long-run relationship between inequality and growth in post-reform China: New evidence from dynamic panel model. *International Review of Economics and Finance*, In press.
- Marriott, F.H.C. and Pope, J.A. 1954. "Bias in the estimation of autocorrelations." *Biometrika*, Vol. 41, pp. 390-402.
- Patterson, K. 2000. "Finite sample bias of the least squares estimators in an AR(p) model: estimation, inference, simulation and examples." *Applied Economics*, Vol. 32 No. 15, pp. 1993-2005.
- Phillips, P.C.B. 1977. "Approximations to Some Finite Sample Distributions Associated with a First-Order Stochastic Difference Equation." *Econometrica*, Vol. 45, No. 2, pp. 463-485.
- Teulings, C., & Van Rens, T. (2008). Education, growth, and income inequality. *The Review of Economics and Statistics*, Vol. 90(1), pp. 89-104.

- Wilkins, A. 2015. To Lag or Not to Lag? Re-evaluating the Use of Lagged Dependent Variables in Regression Analysis. Working paper, Department of Political Science, Stanford University.
- Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach* (5th ed.). Mason, Ohio: South-Western Cengage Learning.

TABLE 1
Sample of Studies that Estimate $LRP = \hat{\beta}_x / (1 - \hat{\beta}_y)$

<i>Author(s)</i>	<i>Journal (Year, Page¹)</i>	<i>Data and Estimation Method</i>
Bentzen & Engsted	<i>Energy</i> (2001, p. 47)	Panels = No Time Period = 1960–1996 Estimation Method = OLS $\hat{\beta}_y$ estimate: 0.657
Carstensen & Toubal	<i>Journal of Comparative Economics</i> (2004, p. 11)	Panels = 60 Time Period = 1993–1999 Estimation Method = DPD $\hat{\beta}_y$ estimates: 0.198 to 0.348
Horioka & Wan	<i>Journal of Money, Credit and Banking</i> (2007, p. 2088)	Panels = 31 Time Period = 1996–2004 Estimation Method = DPD $\hat{\beta}_y$ estimates: 0.476 to 0.628
Dergiades & Tsoulfidis	<i>Energy Economics</i> (2008, p. 2725)	Panels = No Time Period = 1965 to 2006 Estimation Method = OLS $\hat{\beta}_y$ estimates: Not reported

<i>Author(s)</i>	<i>Journal (Year, Page¹)</i>	<i>Data and Estimation Method</i>
Teulings & Van Rens	<i>Review of Economics and Statistics</i> (2008, p. 95)	Panels1 = 24 Obs/Panel1 = 2.3 Panels2 = 95-111 Obs/Panel2 $\approx$ 2.6-3.6 Estimation Method = OLS, WLS, DPD $\hat{\beta}_y$ estimates: 0.374 to 0.955 (Table 3); 0.230 to 0.946 (Table 4)
Baltagi, Demetriades & Law	<i>Journal of Development Economics</i> (2009, p. 287)	Panels = 21-42 Time Observations = 14-22 Estimation Method = DPD $\hat{\beta}_y$ estimates: 0.829 to 0.928 (Table 3); 0.701 to 0.932 (Table 4); 0.716 to 0.792 (Table 5); 0.412 to 0.837 (page 294)
Hoque & Yusop	<i>Journal of Asian Economics</i> (2010, p. 43)	Panels = No Time Period = 1972/1973 – 2004/2005 Estimation Method = OLS $\hat{\beta}_y$ estimates: Not reported
Heid, Langer & Larch	<i>Economics Letters</i> (2012, p. 167)	Panels = 92-150 Obs/Panel1 = 6.28-7.23 Estimation Method = OLS, DPD $\hat{\beta}_y$ estimates: 0.379 to 0.735

<i>Author(s)</i>	<i>Journal (Year, Page¹)</i>	<i>Data and Estimation Method</i>
Islam, Shahbaz, Ahmed & Alam	<i>Economic Modelling</i> (2013, p. 438)	Panels = No Time Period = 1971–2009 Estimation Method = OLS $\hat{\beta}_y$ estimates: Not reported
De Vita	<i>Tourism Management</i> (2014, p. 231)	Panels $\approx$ 295 Time Period = 1980-2004/2011 Estimation Method = DPD $\hat{\beta}_y$ estimates: Not reported
Li, Lai, Wang & Zhao	<i>International Review of Economics & Finance</i> (2015, p. 4) http://dx.doi.org/10.1016/j.iref.2015.08.009	Panels $\approx$ 27 Time Period $\approx$ 1984-2012 Estimation Method = DPD $\hat{\beta}_y$ estimates: Not reported

¹ Page number identifies the page where the article explicitly states that it estimates the long-run effect by $\hat{\beta}_x / (1 - \hat{\beta}_y)$.

TABLE 2
Demonstration of Finite-Sample Bias in OLS Estimation of an AR Model

<i>True Value of β_y</i>	<i>SAMPLE SIZE</i>				
	<i>T=15</i>	<i>T=30</i>	<i>T=50</i>	<i>T=100</i>	<i>T=1000</i>
<i>Average Estimated Value of β_y</i>					
<i>0.10</i>	0.014	0.057	0.074	0.086	0.098
<i>0.50</i>	0.339	0.418	0.449	0.475	0.498
<i>0.90</i>	0.640	0.764	0.819	0.860	0.896
<i>0.95</i>	0.668	0.802	0.862	0.906	0.946
<i>0.99</i>	0.698	0.830	0.893	0.941	0.986
<i>Type I Error Rate ($H_0: \beta_y = \text{true value}$)</i>					
<i>0.10</i>	0.041	0.046	0.051	0.047	0.054
<i>0.50</i>	0.052	0.054	0.054	0.051	0.053
<i>0.90</i>	0.131	0.129	0.103	0.085	0.055
<i>0.95</i>	0.178	0.165	0.144	0.110	0.056
<i>0.99</i>	0.221	0.238	0.227	0.203	0.089

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 2 of the text.

TABLE 3
Estimation of *LRP* in an ARDL Model: OLS

<i>True Value of β_y / True Value of LRP of x</i>	<i>SAMPLE SIZE</i>		
	<i>T=10</i>	<i>T=50</i>	<i>T=1000</i>
<i>Median Value of Estimated LRP of x</i>			
<i>0.10 / 1.11</i>	1.07	1.11	1.11
<i>0.50 / 2</i>	1.6	1.9	2.0
<i>0.90 / 10</i>	2.8	7.5	9.9
<i>0.95 / 20</i>	2.8	11.4	19.5
<i>0.99 / 100</i>	2.7	16.0	85.9
<i>90 Percent Empirical Sample Range for Estimated LRP of x</i>			
<i>0.10 / 1.11</i>	0.3 — 2.3	0.8 — 1.5	1.0 — 1.2
<i>0.50 / 2</i>	0.4 — 5.0	1.3 — 2.8	1.8 — 2.2
<i>0.90 / 10</i>	-13.2 — 21.9	3.4 — 19.4	8.3 — 11.7
<i>0.95 / 20</i>	-18.1 — 25.2	3.9 — 45.7	15.3 — 24.8
<i>0.99 / 100</i>	-24.6 — 28.8	-72.1 — 125.4	51.2 — 152.9
<i>Type I Error Rate (H_0: LRP of x = true value)</i>			
<i>0.10 / 1.11</i>	0.104	0.064	0.050
<i>0.50 / 2</i>	0.185	0.081	0.051
<i>0.90 / 10</i>	0.530	0.267	0.069
<i>0.95 / 20</i>	0.661	0.393	0.084
<i>0.99 / 100</i>	0.844	0.681	0.197

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 3 of the text.

TABLE 4
Estimation of LRP in a DPD Model ($N=50$, $T=10$): AH, DGMM, and SGMM

<i>True Value of β_y / True Value of LRP of x</i>	<i>Anderson-Hsaio</i>	<i>ESTIMATOR Difference GMM</i>	<i>System GMM</i>
<i>Median Value of Estimated LRP of x</i>			
<i>0.10 / 1.11</i>	1.11	1.08	1.11
<i>0.50 / 2</i>	1.9	1.8	2.0
<i>0.90 / 10</i>	0.6	5.1	25.6
<i>0.95 / 20</i>	2.5	10.8	-56.1
<i>0.99 / 100</i>	8.3	34.7	-22.2
<i>Average 90 Percent Empirical Sample Range for Estimated LRP of x</i>			
<i>0.10 / 1.11</i>	0.9 — 1.3	1.0 — 1.2	1.0 — 1.2
<i>0.50 / 2</i>	1.3 — 3.5	1.6 — 2.1	1.7 — 2.3
<i>0.90 / 10</i>	-5.9 — 7.0	3.3 — 8.3	11.4 — 99.7
<i>0.95 / 20</i>	-27.3 — 27.8	6.6 — 21.1	-332.5 — 225.6
<i>0.99 / 100</i>	-99.4 — 105.0	13.1 — 170.0	-30.0 — -17.9
<i>Average Type I Error Rate (H_0: LRP of x = true value)</i>			
<i>0.10 / 1.11</i>	0.053	0.083	0.088
<i>0.50 / 2</i>	0.077	0.194	0.110
<i>0.90 / 10</i>	0.709	0.821	0.001
<i>0.95 / 20</i>	0.514	0.642	0.499
<i>0.99 / 100</i>	0.524	0.585	1.000

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 4 of the text. AH, DGMM, and SGMM stand for Anderson-Hsaio, Difference GMM, and System GMM.

TABLE 5
Estimation of *LRP* in a DPD Model ($N=140$, $T=5$): AH, DGMM, and SGMM

<i>True Value of β_y / True Value of <i>LRP</i> of x</i>	<i>Anderson-Hsaio</i>	<i>ESTIMATOR Difference GMM</i>	<i>System GMM</i>
<i>Median Value of Estimated <i>LRP</i> of x</i>			
<i>0.10 / 1.11</i>	1.11	1.09	1.10
<i>0.50 / 2</i>	2.0	1.8	1.9
<i>0.90 / 10</i>	0.8	3.4	25.2
<i>0.95 / 20</i>	1.8	7.7	-57.2
<i>0.99 / 100</i>	8.3	25.6	-22.1
<i>Average 90 Percent Empirical Sample Range for Estimated <i>LRP</i> of x</i>			
<i>0.10 / 1.11</i>	0.9 — 1.3	1.0 — 1.2	1.0 — 1.2
<i>0.50 / 2</i>	1.4 — 3.3	1.6 — 2.2	1.7 — 2.3
<i>0.90 / 10</i>	-8.7 — 9.6	2.1 — 6.2	10.8 — 105.2
<i>0.95 / 20</i>	-22.0 — 21.5	4.5 — 19.1	-263.0 — -21.4
<i>0.99 / 100</i>	-103.1 — 105.6	-86.0 — 148.9	-27.2 — -18.6
<i>Average Type I Error Rate (H_0: <i>LRP</i> of x = true value)</i>			
<i>0.10 / 1.11</i>	0.052	0.083	0.095
<i>0.50 / 2</i>	0.073	0.191	0.143
<i>0.90 / 10</i>	0.640	0.921	0.001
<i>0.95 / 20</i>	0.583	0.746	0.618
<i>0.99 / 100</i>	0.540	0.618	1.000

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 4 of the text. AH, DGMM, and SGMM stand for Anderson-Hsaio, Difference GMM, and System GMM.

TABLE 6
Comparison of Jackknifing and Indirect Inference with OLS Estimates of β_y in an ARDL Model

<i>True Value of β_y</i>	<i>SAMPLE SIZE</i>				
	<i>T=10</i>	<i>T=30</i>	<i>T=50</i>	<i>T=100</i>	<i>T=1000</i>
<i>OLS estimate of β_y</i>					
<i>0.10</i>	0.03	0.08	0.09	0.09	0.10
<i>0.50</i>	0.37	0.46	0.48	0.49	0.50
<i>0.90</i>	0.69	0.83	0.86	0.88	0.90
<i>0.95</i>	0.73	0.87	0.90	0.93	0.95
<i>0.99</i>	0.76	0.91	0.94	0.96	0.99
<i>Jackknifing estimate of β_y</i>					
<i>0.10</i>	0.09	0.10	0.10	0.10	0.10
<i>0.50</i>	0.50	0.50	0.50	0.50	0.50
<i>0.90</i>	0.85	0.90	0.90	0.90	0.90
<i>0.95</i>	0.89	0.95	0.95	0.95	0.95
<i>0.99</i>	0.91	0.98	0.99	0.99	0.99
<i>Indirect Inference estimate of β_y</i>					
<i>0.10</i>	0.10	0.10	0.10	0.10	0.10
<i>0.50</i>	0.50	0.50	0.50	0.50	0.50
<i>0.90</i>	0.86	0.90	0.90	0.90	0.90
<i>0.95</i>	0.89	0.94	0.95	0.95	0.95
<i>0.99</i>	0.92	0.97	0.98	0.99	0.99

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 5 of the text.

TABLE 7
Estimation of *LRP* in an ARDL Model: Jackknifing and Indirect Inference

<i>True Value of β_y / True Value of <i>LRP</i> of x</i>	<i>T = 10</i>		<i>T = 50</i>		<i>T = 1000</i>	
	<i>JK</i>	<i>II</i>	<i>JK</i>	<i>II</i>	<i>JK</i>	<i>II</i>
<i>Median Value of Estimated <i>LRP</i> of x</i>						
<i>0.10 / 1.11</i>	1.05	1.14	1.11	1.11	1.11	1.11
<i>0.50 / 2</i>	1.5	1.9	2.0	2.0	2.0	2.0
<i>0.90 / 10</i>	1.0	2.0	8.2	9.9	10.1	10.1
<i>0.95 / 20</i>	0.6	1.4	9.3	12.8	20.3	20.3
<i>0.99 / 100</i>	0.5	0.8	6.8	8.4	100.1	105.1
<i>90 Percent Empirical Sample Range for Estimated <i>LRP</i> of x</i>						
<i>0.10 / 1.11</i>	-0.2 – 3.7	0.3 - 2.7	0.8 – 1.5	0.8 – 1.5	1.0 – 1.2	1.0 – 1.2
<i>0.50 / 2</i>	-6.6 – 10.2	-0.7 – 10.1	1.3 – 3.2	1.3 – 3.0	1.8 – 2.2	1.8 – 2.2
<i>0.90 / 10</i>	-18.7 – 19.0	-34.7 – 34.2	-35.1 – 52.0	-36.5 – 55.2	8.4 – 12.0	8.4 – 12.0
<i>0.95 / 20</i>	-18.6 – 19.6	-37.5 – 35.5	-77.4 – 85.1	-128.9 – 127.5	15.7 – 26.6	15.8 – 26.0
<i>0.99 / 100</i>	-18.9 – 18.8	-38.1 – 33.4	-98.2 – 103.0	-184.6 – 186.4	49.2 – 330.1	56.8 – 249.4
<i>Average Type I Error Rate (H_0: <i>LRP</i> of x = true value)</i>						
<i>0.10 / 1.11</i>	0.231	0.093	0.074	0.058	0.051	0.051
<i>0.50 / 2</i>	0.278	0.131	0.088	0.059	0.054	0.053
<i>0.90 / 10</i>	0.419	0.392	0.252	0.133	0.062	0.055
<i>0.95 / 20</i>	0.543	0.541	0.288	0.193	0.075	0.055
<i>0.99 / 100</i>	0.783	0.787	0.509	0.514	0.201	0.095

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 5 of the text.

TABLE 8
Estimation of LRP in a DPD Model ($N=50, T=10$): Jackknifing and Indirect Inference

<i>True Value of β_y / True Value of LRP of x</i>	<i>Anderson-Hsaio</i>		<i>Difference GMM</i>		<i>System GMM</i>	
	<i>JK</i>	<i>II</i>	<i>JK</i>	<i>II</i>	<i>JK</i>	<i>II</i>
<i>Median Value of Estimated LRP of x</i>						
<i>0.10 / 1.11</i>	1.11	1.11	1.11	1.11	1.11	1.10
<i>0.50 / 2</i>	2.0	2.0	2.0	2.0	2.0	2.0
<i>0.90 / 10</i>	0.4	0.6	8.3	7.3	13.3	12.9
<i>0.95 / 20</i>	0.1	1.2	16.7	-0.9	57.6	60.3
<i>0.99 / 100</i>	4.5	9.6	-31.6	11.2	-38.1	-40.3
<i>90 Percent Empirical Sample Range for Estimated LRP of x</i>						
<i>0.10 / 1.11</i>	0.9 — 1.4	0.7 – 1.7	1.0 – 1.2	0.9 – 1.5	1.0 – 1.3	0.9 – 1.4
<i>0.50 / 2</i>	1.3 — 4.0	1.1 – 4.6	1.7 – 2.4	0.8 – 6.6	1.7 – 2.4	1.0 – 5.6
<i>0.90 / 10</i>	-9.1 — 9.7	-18.6 – 20.2	4.3 – 30.0	-60.2 – 55.3	7.5 – 35.1	-90.2 – 105.3
<i>0.95 / 20</i>	-6.7 — 6.3	-51.6 – 60.3	-98.8 – 137.8	-460.9 – 260.3	-456.6 – 508.7	-220.3 – 390.3
<i>0.99 / 100</i>	-78.4 — 76.3	-103.1 – 127.3	-333.4 – 330.5	-263.1 – 318.3	-56.1 – -29.5	-663.1 – 18.3
<i>Average Type I Error Rate (H_0: LRP of x = true value)</i>						
<i>0.10 / 1.11</i>	0.296	0.199	0.291	0.139	0.278	0.211
<i>0.50 / 2</i>	0.337	0.188	0.258	0.200	0.255	0.189
<i>0.90 / 10</i>	0.158	0.362	0.213	0.663	0.524	0.888
<i>0.95 / 20</i>	0.775	0.133	0.780	0.863	0.435	0.536
<i>0.99 / 100</i>	0.484	0.659	0.270	0.370	1.000	1.000

SOURCE: 10,000 replications were run for each experiment. The experiments are described in more detail in Section 5 of the text.