Containing Our Confidence: Controlling Explosive Confidence Intervals when using Long Run Multipliers

Mark David Nieman* and David A.M. Peterson†

Abstract

The recent exchange on Error Correction Models in *Political Analysis* and elsewhere dealt with several important issues involved in time series analysis. While there was much disagreement in the symposium, one common theme was the lack of power due to the few number of observations for much of this work. As is well known, one result of low power is inflated standard errors. One issue low powered time series often face is that the confidence interval on a lagged dependent variable, even when the series is stationary, includes values ≥ 1. This is particularly problematic when calculating the confidence interval of the long run multiplier. If the confidence interval of the lagged dependent variable includes 1, the standard error of the long run multiplier will be explosive. As a solution, we suggest using a Bayesian approach which formalizes the stationarity assumption by using a beta prior that is strictly less than 1. As a result, we obtain theoretically informed estimates of the confidence regions for the lagged of the distribution of the long run multiplier.

*Assistant Professor of Political Science, Iowa State University.
†Professor of Political Science, Iowa State University.

1. INTRODUCTION

The recent exchange on the utility of and concerns with error correction models did not generate much consensus on the appropriateness of those models (e.g., Enns et al, 2016; Grant and Lebo, 2016; Keele, Linn, and Webb, 2016). What is notable, however, that there were at last two places of agreement between the authors that have implications for users of time series models. First, as most clearly articulated by DeBoef and Keele (2008), it is important for practitioners to calculate the long run multiplier (LRM) and its standard error to present a complete picture of the effect of an independent variable. Second, many of our time series applications are short and the small number of observations limits the power of the estimation of specification tests and important parameters.[1] In the case of an autoregressive distributed lag (ADL) model, one potential problem is that the lack of power in the estimation may lead to the estimation of the coefficient for the lagged dependent variable, even if the series is stationary, where the confidence interval includes 1.[2] In that case, the estimation of the variance of the LRM will be "mildly explosive"[3] (which is bad).

In this paper, we examine the effects of having a short series on the estimation of the LRM in particular and suggest a simple fix that will provide more accurate estimates of the uncertainty in the LRM. Specifically, we propose placing a prior on the lagged dependent variable in an ADL model that constrains the coefficient of the lagged dependent variable to be strictly between zero and one. This prior simply incorporates the decision to treat the series as stationary and, as we will show, leads to much more accurate estimates of the uncertainty in the LRM and the specification of the ADL. In the next section, we discuss the general problem, outlining the main methods for calculating the LRM and its uncertainty. We then present a series of Monte Carlos that illustrate the problems with each of theses approaches and the improvements due to the inclusion of a Beta prior on the lagged dependent variable.

2. THE UNCERTAINTY OF THE LRM

The ADL model is standard approach to modeling stationary series.[4] The standard set up for an ADL(1,1;1) is:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t \tag{1}$$

[1] Freeman (2016) provides a counter to this ubiquity of small T time series.

[2] One may suggest that if the confidence interval for α_1 contains 1, then the researcher should not treat the series as stationary. While this might seem like a reasonable suggestion, there may be cases where stationarity is still appropriate for small samples. Undoubtedly, if a the coefficient on the lagged dependent variable is not significantly different than 1, that series is unlikely to reject the null of a unit root in stationarity tests. Researchers, however, often have more knowledge about the properties of their dependent variable beyond the small set of observations used. Researchers who are limited by the timespan of their independent variables, for instance, may have a series that is stationary with larger T than they can use in their analyses. The small T also means that that the stationarity tests themselves are often weak.

[3] We borrow this phrase from Hill and Peng (2014, 293) and Hill, Li, and Peng (2016, 126).

[4] We explicitly assume that the researcher is treating the dependent variable as stationary. That is, we set aside all concerns about the small sample properties of stationarity tests, and issues of fractional or near integration. If the researcher is not going to treat the dependent variable as stationary, none of the issues we discuss here necessarily apply.

In this design, the LRM is calculated as $lrm_{\text{ADL}} = \frac{\beta_0+\beta_1}{1-\alpha_1}$ and captures the cumulative effect of a one unit change in X_t on Y_t summed across the timespan. As such, it provides a summary of the substantive effect of X. While they are not always reported in times series analysis, they have become more standard in the literature. Unfortunately, the variance associated with the estimate of the LRM seems to be much less likely to be reported.

The absence of these uncertainty estimates is due, in part, to the difficulty in calculating variance of a ratio of parameters. In fact, existing measures of uncertainty are for the LRM are typically calculated from an error correction model (ECM), which DeBoef and Keele (2008: 189-190) have shown can be written as mathematically equivalent to the ADL. Equation 1 in ECM form is:

$$\Delta Y_t = \alpha_0 + (\alpha_1 - 1)Y_{t-1} + \beta_0 \Delta X_t + (\beta_0 + \beta_1)X_{t-1} + \epsilon_t \tag{2}$$

and the LRM is calculated as $lrm_{\text{ECM}} = \frac{\beta_0+\beta_1}{\alpha_1-1}$. As DeBoef and Keele (2008: 191) note, neither the ADL nor the ECM provide a direct estimate of the LRM. Instead, since the ECM is a ratio of coefficients, the calculation of the variance of the ratio of coefficients with known variances can be used. The formula is:

$$\text{Var}\left(\frac{a}{b}\right) = \left(\frac{1}{b^2}\right)\text{Var}(a) + \left(\frac{a^2}{b^4}\right)\text{Var}(b) - 2\left(\frac{a}{b^3}\right)\text{Cov}(a,b). \tag{3}$$

An alternative calculation of the LRM from an ECM is to use the Bewley (1979) transformation, which estimates the variance of the LRM directly.[5] The Bewley transformation is:

$$Y_t = \alpha_0\phi - \alpha_1\phi\Delta Y_t + \phi\left(\beta_0 + \beta_1\right)X_t + \phi\beta_1\Delta X_t + \phi\epsilon_t \tag{4}$$

where $\phi = \left(\frac{1}{\alpha_1-1}\right)$ and an instrument for ΔY_t is calculated as the predicted values from the equation $\Delta Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 X_t + \gamma_3 \Delta X_t + \epsilon_t$. The LRM is the coefficient on X_t from equation 4. These two approaches are asymptotically equivalent.

In small samples, however, both of these techniques face difficulties. When the T is small, the estimate of α_1 can be imprecise. Normally, this lack of power will simple lead to inflated standard errors and not many other issues. In this case, however, there is an additional potential pitfall. If the confidence interval of α_1 includes 1, the variance calculation will be "slightly explosive". This can be seen in the calculation of the LRM. If the value of α_1 has probability mass at 1. the LRM is undefined for that point. If there is mass where $\alpha_1 > 1$, the denominator will be negative. In each case, the calculation of the variance breaks down.

3. BAYESIAN APPROACH

Our approach to this problem is to use a Bayesian framework. This allows us to fully integrate the stationarity of the dependent variable by adding a prior on the coefficient that will constrain it to be strictly between 0 and 1. In the example presented here, we use the prior such that $\alpha_1 \sim \mathcal{B}(1,1)$ to accomplish this. This is a diffuse prior that places equal

[5]The Bewley transformation is a computational convenience and is not a theoretical model.

probability on all values between 0 and 1. A more informed prior, which places more weight to values closer to 1 may be practical, but for this example we illustrate the influence of this design with a flat prior.[6]

One way to think about this prior is that it is simply the formalization of stationarity. When a researcher treats a series a stationary, she is assume that the root of the characteristic equation of the time series is less than one. The use of this prior constrains the estimate of α_1 to be less than 1, precisely the implication of treating the dependent variable as stationary.

The actual estimation of the model is carried out via a Markov chain Monte Carlo (MCMC). This allows us to calculate the distribution of the posterior for all of the coefficients directly. Rather than using an asymptotic equivalent to the confidence interval of the LRM or relying on a formula that is not easily available in most statistical output, we can calculate the LRM for each of the draws from the posterior in the MCMC and using this distribution to summarize the credible region of the LRM.[7]

4. MONTE CARLO ANALYSIS

To compare the properties of this approach versus the standard techniques, we conduct two Monte Carlos. We generate data following an ADL(1,1;1). To explore the small sample properties, we set $T = 50$. We generate the exogenous variable so that $X_t = \gamma X_{t-1} + \eta_t$ where $\gamma = 0.5$.[8] We generate the endogenous variable so that $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$ where $\alpha_0 = 0$, $\beta_0 = 0.5$, and $\beta_1 = 0.25$. η_t and ϵ_t are drawn from a standard normal where $\text{cov}(\eta_t, \epsilon_t = 0)$. We vary $\alpha_1 = \{0.5, 0.9\}$. In the case where $\alpha_1 = 0.5$, the LRM is $\frac{0.5+0.25}{1-0.5} = 1.5$; when $\alpha_1 = 0.5$, the LRM is $\frac{0.5+0.25}{1-0.9} = 7.5$.

We estimate an ADL, ECM, and Bayesian ADL (B-ADL) model.[9] Priors for the Bayesian ADL are: $\beta_j \sim \mathcal{N}(0, 20)$, $\alpha_0 \sim \mathcal{N}(0, 20)$, $\alpha_1 \sim \mathcal{B}(1, 1)$, and $\epsilon \sim \mathcal{G}(1, 10)$. The Beta prior has a uniform probability mass function (PMF) strictly between 0 and 1. Each Bayesian ADL uses 10,000 MCMCs after a 2,500 burnin. Estimates are based on 250 simulations for each value of α_1.

5. RESULTS

Table 1 reports the results from the Monte Carlos. Each of the models recovers the expected coefficients when $\alpha_1 = 0.5$. Moreover, the LRM is also correctly estimated by each of the models. When $\alpha_1 = 0.9$, we again find that each of the models is able to recover the correct coefficients. The LRM is the same for the ADL and ECM (both using the calculation of the variance of the ratio of coefficients, or *ECM-ratio*, and the Bewley transformation, or *ECM-Bewley*), while the B-ADL somewhat underestimates the LRM. The latter result stems

[6]If one has theoretical reasons to expect the coefficient on the lagged dependent variable to approach 1, a Beta prior assuming a pmf massed near 1, such as $\mathcal{B}(5, 2)$, could be used. For the sake of demonstrating the approach, however, we use $\mathcal{B}(1, 1)$, which will calculate a more conservative LRM.

[7]Implementing this approach can be easily accomplished in common statistical programs, such as R or Stata.

[8]The initial X is drawn from a standard normal distribution.

[9]The LRM for the B-ADL is calculated by recovering the median posterior. We also calculate the LRM for the ECM using the Bewley transformation.

Table 1: B-ADL, ADL, ECM Estimates.

	$\alpha_1 = 0.5$				$\alpha_1 = 0.9$		
Variable	B-ADL	ADL	ECM	Variable	B-ADL	ADL	ECM
Y_{t-1}	0.46	0.46	-0.54	Y_{t-1}	0.87	0.87	-0.13
	(0.11)	(0.11)	(0.11)		(0.05)	(0.05)	(0.05)
X_t	0.49	0.49		X_t	0.48	0.48	
	(0.17)	(0.17)			(0.17)	(0.17)	
X_{t-1}	0.28	0.28	0.77	X_{t-1}	0.29	0.28	0.77
	(0.20)	(0.19)	(0.16)		(0.17)	(0.17)	(0.13)
ΔX_t			0.49	ΔX_t			0.48
			(0.17)				(0.17)
Const.	0.00	0.00	0.00	Const.	0.03	0.03	0.03
	(0.15)	(0.15)	(0.15)		(0.16)	(0.16)	(0.16)
LRM	1.47	1.47	1.47	LRM	7.08	7.29	7.29
	(0.81)		(0.29)		(3.41)		(5.53)
T	49	49	49	T	49	49	49

Note: Bewley LRM($\beta_{\alpha_1=0.5}$)=1.47, LRM(SE$_{\alpha_1=0.5}$)=0.15 and Bewley LRM($\beta_{\alpha_1=0.5}$)=7.29, LRM(SE$_{\alpha_1=0.5}$)=0.39. B-ADL LRM: HDP 95%CI$_{\alpha_1=0.5}$=[0.89, 2.20], HDP 95%CI$_{\alpha_1=0.9}$=[2.27,32.02].

from the inclusion of the flat prior in the lower power context. Overall, each of the models performs well in terms of bias.

Turning to the estimates of uncertainty, we find that the standard errors on the coefficients are consistent across models for both $\alpha_1 = 0.5$ and $\alpha_1 = 0.9$. With small sample sizes, none of the models indicate that β_1 is statistically significant, even though it is a predictor of Y_t in the true data generating process. This is potentially problematic, as applied researchers often disregard insignificant lags of independent variables, which would result in biased estimates of the LRM.

Figure 1: Standard Error on Long Run Multiplier

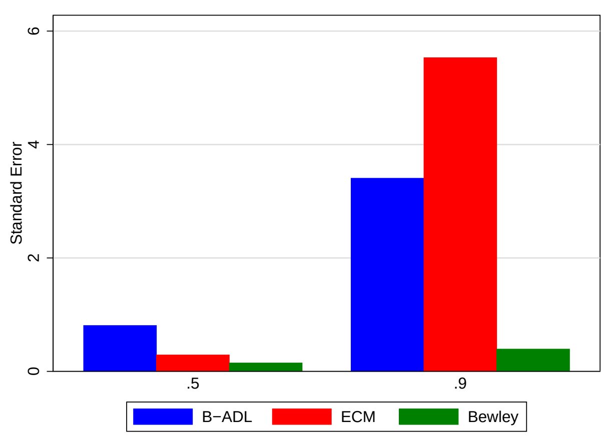
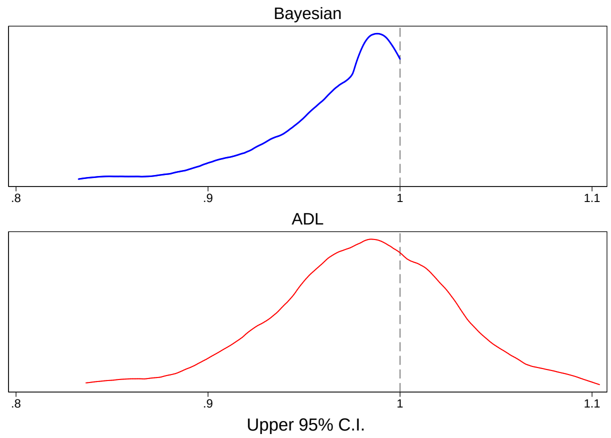


Figure 2: Upper C.I. on Lagged DV when $\alpha_1 = 0.9$



Note: 1 is the theoretical upper bound.

Uncertainty estimates for the LRM differ substantially across models. Figure 1 report the estimated standard errors for the B-ADL, ECM-ratio, and ECM-Bewley for $\alpha_1 = 0.5$ and $\alpha_1 = 0.9$. When $\alpha_1 = 0.5$, the B-ADL model has a SE that over twice as large as the ECM-ratio, and the ECM-ratio is almost twice as large as ECM-Bewley. The former is accounted for by the influence of the prior given the small sample size. The latter result reflects the fact that ECM-Bewley is an instrumental approach and thus tends to underestimate the level of uncertainty in a small sample setting. The difference between the ECM-ratio and ECM-Bewley techniques demonstrates that, while they recover the same asymptotic results, they differ in small samples.

When $\alpha_1 = 0.9$, differences in the estimates of the SE are even more stark. SEs for the LRM are approximately half the size of the coefficient for B-ADL, while the SEs for ADL are over 5.5, reflecting the fact that ADL contains some simulations where the confidence interval are "mildly explosive," or greater than 1. Figure 2 makes this point more clear. The prior on B-ADL forces estimates of the standard errors to remain within the range of theoretically possible outcomes, whereas a non-insignificant proportion of the upper 95% confidence intervals for ADL are greater than 1. In contrast, the SEs for ECM-Bewley are much smaller than those of either the B-ADL or ADL models.

6. CONCLUSION

Our results demonstrate that when the coefficient on the lagged DV is moderate to small, each approach is able to recover correct parameter estimates, including for the LRM. When the coefficient on the lagged DV is large, however, B-ADL is preferable, as it is the only approach that provides theoretically appropriate estimates of the uncertainty on the LRM; ECM-ratio suffers from "explosive" SEs and ECM-Bewley underestimates the degree of uncertainty when sample sizes are small.

In addition, the results reveal that, in small samples, analysts may underestimate the LRM because the coefficient for X_{t-1} is not statistically significant in any of the models at either value of α. This suggests that applied researchers should be especially careful when selecting the lag structure of the independent variables, as the coefficients on such variables may be incorrectly identified as lacking statistical significance owing to low power. Discarding such variables, of course, would lead to inaccurate estimates of the LRM.

REFERENCES

Bewley, Ronald A. 1979. The direct estimation of the equilibrium response in a linear dynamic model. *Economics Letters* 3(4): 357-361.

De Boef, Suzanna and Luke Keele. 2008. Taking Time Seriously. *American Journal of Political Science* 52(1): 184-200.

Enes, Peter K., Nathan J. Kelly, Takaaki Masaki, and Patrick C. Wohlfarth. 2016. Don't Jettison the General Error Correction Model Just Yet: A Practical Guide to Avoiding Spurious Regression with the GECM. *Research and Politics* 3(2): 10.1177/2053168016643345.

Freeman, John R. 2016. Progress in the Study of Nonstationary Political Time Series: A Comment. *Political Analysis* 24(1): 42-49.

Grant, Taylor and Matthew J. Lebo. 2016. Error Correction Methods with Political Time Series. *Political Analysis* 24(1): 3-30.

Hill, Jonathan and Liang Peng. 2014. Unified Interval Estimation for Random Coefficient Autoregressive Models. *Journal of Time Series Analysis* 35(3): 282-297.

Hill, Jonathan, Deyuan Li, and Liang Peng. 2016. Uniform Interval Estimation for an AR(1) Process with AR Errors. *Statistica Sinica* 26(1): 119-136.

Keele, Luke, Suzanna Linn, and Clayton McLaughlin Webb. 2016. Treating Time with All Due Seriousness. *Political Analysis* 24(1): 31-41.