Appendix for:

Detecting True Relationships in Time Series Data with Different Orders of Integration

Peter K. Enns   Carolina Moehlecke   Christopher Wlezien

Contents

Appendix 1 Adding Unrelated I(1) Regressors to a Correctly Specified GECM Does Not Unbalance the Equation

Here, we consider the Kraft, Key, and Lebo analysis where two I(1) variables that are cointegrated are analyzed with unrelated I(1) variables. This presents an interesting scenario because researchers often control for variables that may not be related to the other variables, because omitting a relevant variable can bias results, while including an unrelated variable typically only risks increased standard errors. Enns & Wlezien (2017) argue that if two variables are cointegrated, adding unrelated I(1) variables to the model (while inadvisable) does not affect equation balance because the linear combination of all variables remains unchanged. However, Kraft, Key, and Lebo argue that controlling for unrelated variables creates equation imbalance and thus increases the rate of Type I errors. The declining rate of Type I errors as T increases in Kraft, Key, and Lebo's Figure 2 actually may support what Enns and Wlezien posit; if the Type I error rate in Kraft, Key, and Lebo's Figure resulted because of imbalance, the false positive rate should not decline as the sample size increases. (That is, equation balance does not depend on sample size.) To better understand their results, we repeated Kraft, Key, and Lebo's simulations setting T equal to 5,000 and including 8 unrelated I(1) predictors. The rate of false positives for all unrelated long run multipliers drops to 5.2 percent with a mean value of -.00002. Although we do not recommend adding unrelated I(1) variables to models with cointegrated variables, consistent with Enns and Wlezien, doing so does not appear to create equation imbalance.

Appendix 2 Combined Time Series with Additional Innovation (q)

For the combined time series analysis, the results in the text use the following DGP,

$$Y_t = (x_t^I + x_t^S) \tag{1}$$

We did not add a disturbance to this DGP because the DGP of both x_t^I and x_t^S contain disturbance terms. Nevertheless, we wanted to be sure that our results were not sensitive to this decision. Thus, we conducted the same simulations where the DGP for Y was,

$$Y_t = (x_t^I + x_t^S) + q, q\, N(0,1) \tag{2}$$

The results appear in Table A-1. The pattern of results is strikingly similar to what we saw in Table 1.

Table A-1: Identifying a True Relationship (β_1=1.0) between x_t^S and Y when x_t^S is Stationary and Y Combines Stationary *and* Unit Root Properties (When Y_t contains the additional innovation q.)

	T = 5,000																	
	ADL						GECM						First Difference					
	$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%
$\hat{\alpha}_1$	0.9967	100	0.9967	100	0.9967	100	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
$\hat{\alpha^*}_1$	N/A	N/A	N/A	N/A	N/A	N/A	-0.0033	0	-0.0033	0	-0.0033	0	N/A	N/A	N/A	N/A	N/A	N/A
$\hat{\beta}_1$	1.0011	100	1.0001	100	0.9998	100	1.0011	100	1.0001	100	0.9998	100	1.0011	100	1.0001	100	0.9998	100
$\hat{\beta}_2$	-0.9969	100	-0.9970	100	-0.9967	100	N/A	N/A	N/A	N/A	N/A	N/A	-1.0002	100	-1.0003	100	-1.0001	100
$\hat{\beta^*}_2$	N/A	N/A	N/A	N/A	N/A	N/A	0.0042	1.55	0.003	0.5	0.0031	0.5	N/A	N/A	N/A	N/A	N/A	N/A

Notes: coef represents the mean coefficient estimate across 2,000 simulations. % represents the percent of simulations for which we (correctly) reject the null hypothesis of no relationship. Consistent with expectations, across all models $\hat{\beta}_1 \approx 1.0$. As explained in the text, the other parameter estimates also follow expectations.
ADL: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \delta$
GECM: $\Delta Y_t = \alpha_0 + \alpha_1^* Y_{t-1} + \beta_1 \Delta x_t^s + \beta_2^* x_{t-1}^s + \gamma$, where $\alpha_1^* = \alpha - 1$ and $\beta_2^* = \beta_1 + \beta_2$ from the ADL.
First Difference: $\Delta Y_t = \alpha_0^* + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \varepsilon$, where $\Delta Y_t = Y_t - \alpha_1 Y_{t-1}$ from the ADL.

Appendix 3 Results Based on the ADL with Finite Series

Table A-2 reports the results based on estimating ADL models for each simulated time series where T=50, 100, and 200. Here we see exactly what we expect on average given the construction of the series, once again regardless of the ρ of the stationary component. Interestingly, when we see departures from the mean expected value, the departures for the Y_{t-1} and X_{t-1} are equal and opposite. Ultimately, despite different orders of integration on the two sides of the equation, at least in the DGPs, we are able to correctly identify the relationship between the stationary x_t^S and the combined time series Y (which in theory is integrated), even with fairly finite samples. While reassuring, note that our analyses here are based on estimations that time series researchers would not undertake without first diagnosing the stationarity of the variables, which we consider in the text. We did so here because we know the GDP for our DV and IV and their interrelationships, and that the ADL should capture the relationship on average, per Table 1.

Table A-2: ADL results for T = 50, T = 100, T = 200

	T = 50						T = 100						T = 200					
	$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$		$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%	coef	%
$\hat{\alpha}_1$	0.8973	100	0.8944	100	0.8831	100	0.9474	100	0.9470	100	0.9435	100	0.9733	100	0.9735	100	0.9728	100
$\hat{\beta}_1$	1.0008	100	0.9981	100	1.0050	100	1.0021	100	0.9989	100	0.9948	100	0.9998	100	0.9966	100	1.0008	100
$\hat{\beta}_2$	-0.8897	99.8	-0.8902	99.8	-0.8894	99.8	-0.9509	100	-0.9442	100	-0.9388	100	-0.9749	100	-0.9726	100	-0.9755	100

Notes: Consistent with expectations, across all models $\hat{\beta}_1 \approx 1.0$. As explained in the text, the other parameter estimates also follow expectations.
ADL: $Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \beta_1 x_t^s + \beta_2 x_{t-1}^s + \delta$

References

Enns, Peter K. & Christopher Wlezien. 2017. "Understanding Equation Balance in Time Series Regression." *The Political Methodologist* 24(2):2–12.