Supplemental Information to: "How to Avoid Incorrect Inferences (While Gaining Correct Ones) in Dynamic Models"

Andrew Q. Philips*

Forthcoming at *Political Science Research and Methods*

May 15, 2021

Contents

*andrew.philips@colorado.edu. Assistant Professor, Department of Political Science, University of Colorado Boulder, UCB 333, Boulder, CO 80309-0333.

1 Notes on simulations and additional results

In the simulations in the manuscript, rejection rates of β_1 were determined using Wald 95% confidence intervals. Long-run effects and confidence intervals were calculated using the delta method in the `R` package `nlWaldTest` (Komashko 2016). To avoid problems with starting values (c.f., Philips 2018), a "burn-in" period of $t = 100$ was added to all 2000 simulations across each of the various combinations of T and the parameters.

Below are a series of additional results. These include the following:

- Increasing the error variance in y_t from $\varepsilon_t \sim N(0,1)$ to $\varepsilon_t \sim N(0,5)$ for all scenarios in order to increase the amount of noise.

- Measures of efficiency and bias for the short- and long-run effects for each of the simulations. I show mean square error for the short-run effects and—due to the sometimes extreme and/or skewed effects given a non-linear combination of parameter estimates—median square error for the long-run effects. For the spurious relation scenarios (I through IV), these are the squared error around zero. For the related scenarios (V and VI), these are the squared error around the true short-run and long-run effects. For both scenarios, the true short-run effect is always $\beta_1 = 2$; the true long-run effect differs in each scenario based on the value of α (dependence in y_t) and β_2 (coefficient on the lag of x_t). Lower values of mean/median square error indicate a more preferred model, since bias and/or efficiency is lower.

- Power (rejection rates of the effects that $H_0 = 0$) for the non-spurious scenarios (V and VI).

- Box-plots showing the estimated short- and long-run effects across all 2000 simulations for the non-spurious scenarios (V and VI).

To make these additional results as accessible as possible to readers, at the end of each section is a summary paragraph listing the main findings from these additional results.

1.1 Scenario I: $Y_t \sim I(0)$, $X_t \sim I(0)$, and unrelated

In Figure 1 I show the proportion of times that the constructed 95 percent confidence intervals of the short-run effects fail to overlap zero for two I(0), unrelated series. These results increase the error variance in y_t to $\sigma_\varepsilon^2 = 5$. The results are quite similar to those in the main paper.

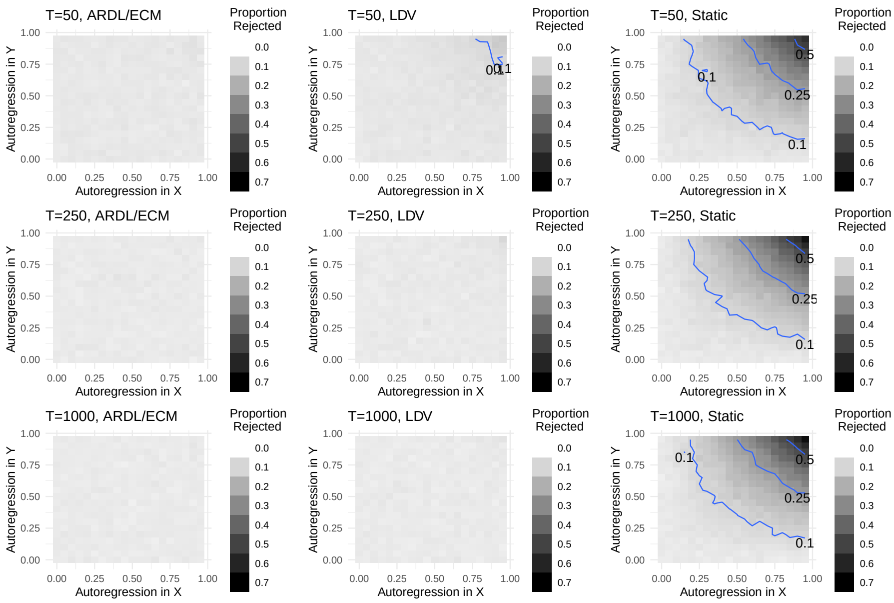

Figure 1: Short-run Type I error, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

Note: Contour lines show boundary of 10, 25, and 50 percent rejection rates.

In Figure 2 I show the proportion of times that the constructed 95 percent confidence intervals of the long-run effects fail to overlap zero for two I(0), unrelated series. These results increase the error variance in y_t to $\sigma_\varepsilon^2 = 5$. Similar to Figure 1, these results are also quite similar to those in the main paper, although the rejection rates appear to be slightly larger.

In Figure 3 I show mean square error (MSE) results for the short-run effect for two I(0), unrelated series. Since the two series are unrelated, the "true" effect is zero. A few things stand out. First, MSE is negligible in the ARDL/ECM and LDV models in larger T. Second, MSE, although higher in the static model, only really appears to be a problem when autoregression is high in the dependent variable. Autoregression in the independent variable appears to be less of an issue, unless y_t is also autoregressive.

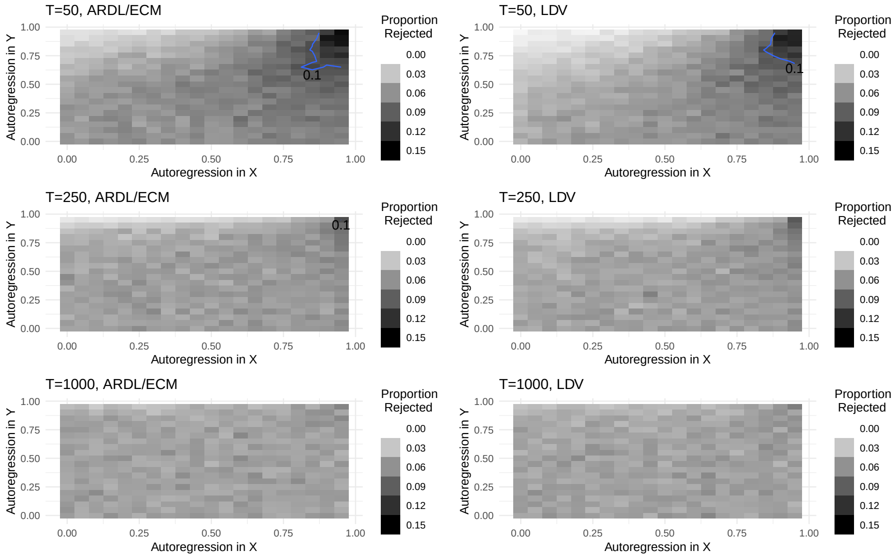

Figure 2: Long-run Type I error, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

Note: Contour lines show boundary of 10 percent rejection rates.

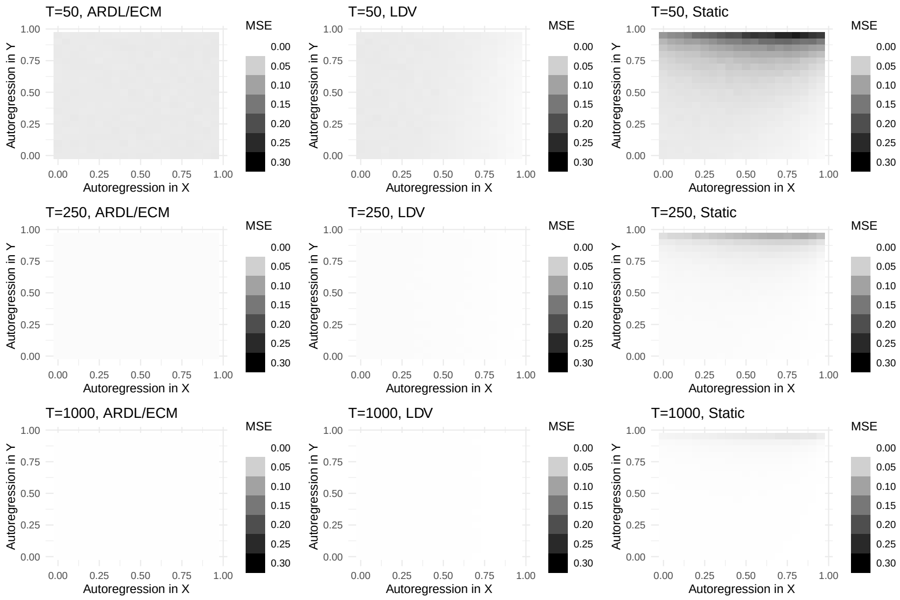

Figure 3: Short-run mean square error, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 1$

Figure 4 shows the same short-run MSE results as Figure 3, but now where $\sigma_\varepsilon^2 = 5$. The results are virtually identical to those in Figure 3, albeit MSE is uniformly larger across all scenarios.

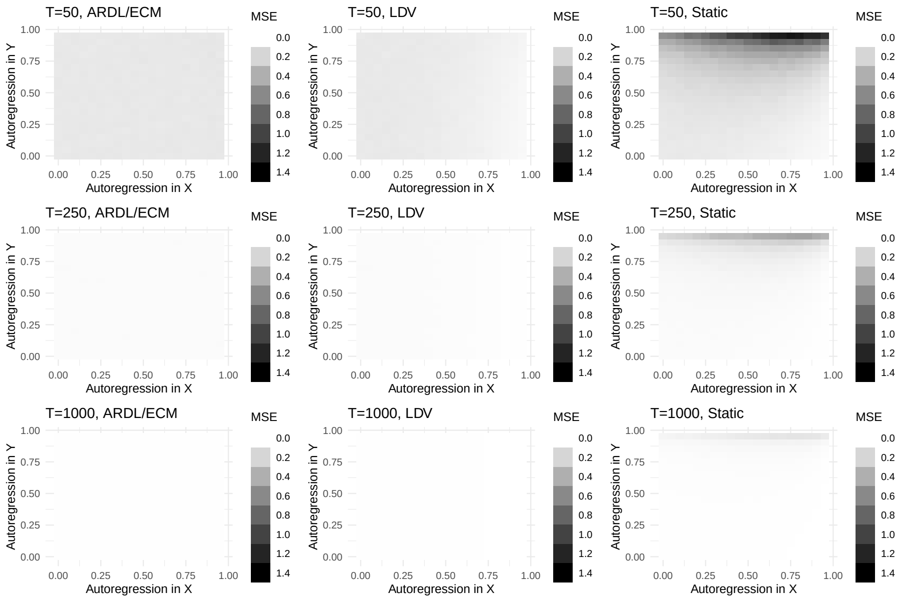

Figure 4: Short-run mean square error, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

In Figures 5 and 6 I show median-square error results for the long-run effect. Recall that long-run effects do not exist for the static model. Figure 5 shows the results when $\sigma_\varepsilon^2 = 1$, while $\sigma_\varepsilon^2 = 5$ in Figure 6. Median square error grows larger as autoregression in the dependent variable approaches one, and tends to be largest when autoregression in the independent variable is low. Across both the ARDL/ECM and LDV model specifications, however, median square error declines as T increases. Figure 6, showing the increased error rates, is similar to Figure 5, although errors are substantially larger, ranging from 0 to 7 in the former and from 0 to 1.5 in the latter.

Summary: Rejection rates for the short-run effect when two series are stationary are only an issue in the static model and get worse as autoregression increases. Rejection rates for the long-run effect are only an issue in the LDV model at extremely high levels of autoregression. MSE appears to only be large in short T, or when autoregression in y_t is high (for the static model in the short-run and ARDL/ECM and LDV in the long-run). High autoregression in y_t appears to affect MSE much more than high autoregression in x_t.

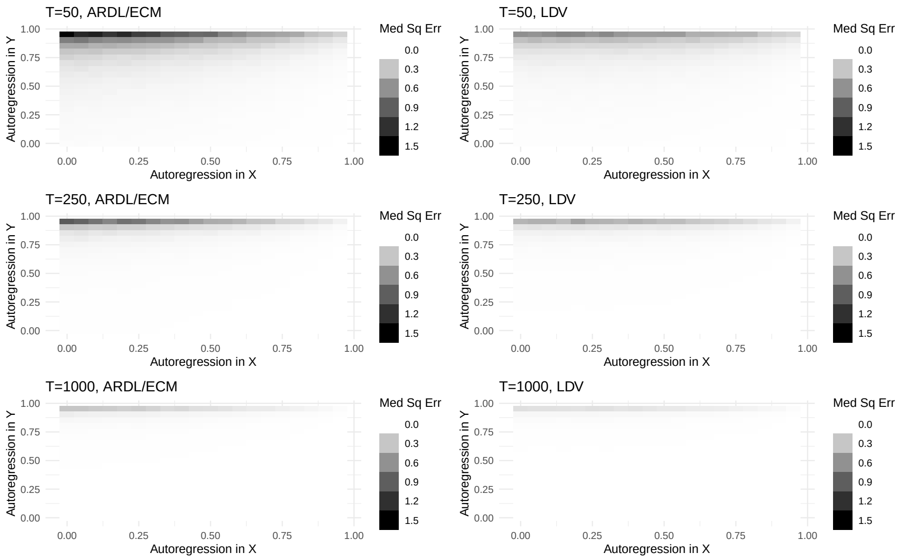

Figure 5: Long-run median square error, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 1$

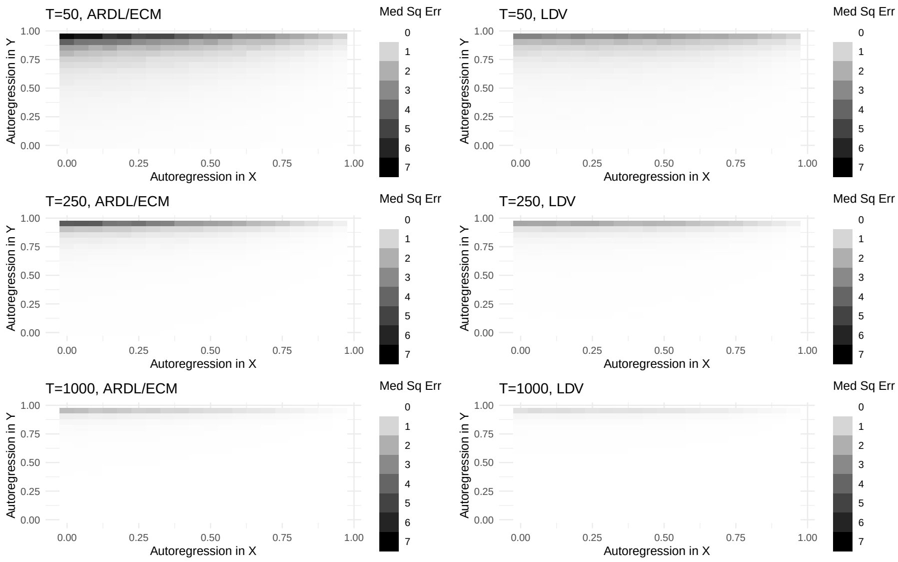

Figure 6: Long-run median square error, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

1.2 Scenario II: $Y_t \sim I(0)$, $X_t \sim I(1)$, and unrelated

In Scenario II, the dependent variable is stationary, yet autoregressive, while the independent variable contains a unit root. Both series are unrelated in this scenario. In Figure 7, I show the short-run rejection rates of $H_0 = 0$, now increasing the error variance of y_t to $\sigma^2_\varepsilon = 5$. These rejection rates are essentially the same—perhaps even slightly lower, than those in the main paper where $\sigma^2_\varepsilon = 1$; throughout all combinations of T, the ARDL/ECM specification seems to perform marginally better than the LDV, although rejection rates for the long-run effect for these three models are higher than convention when $T = 50, 250$ and autoregression in y_t is large.

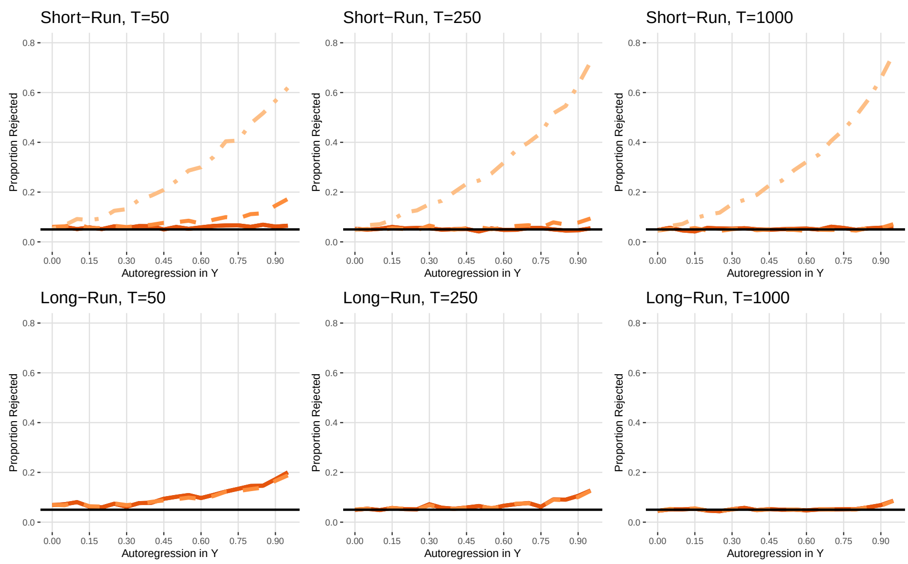

Figure 7: Scenario II: Rejection rates for $Y_t \sim I(0)$, $X_t \sim I(1)$ and $\sigma^2_\varepsilon = 5$

Note: Static (dot-dash), LDV (dash), ARDL/ECM (solid). Long-run effects do not exist for static model.

In Figure 8 I show mean square error for the short-run effects (top row) and median square error for the long-run effects (bottom row). For the short-run, MSE tends to be lowest for the LDV model and highest in the static model when autoregression in y_t is large, although MSE is negligible for all models in large sample sizes. In the long-run, MSE is low when autoregression in y_t is low but grows as it increases, no matter whether the ECM, ARDL, or LDV model is used. This is only a problem in short samples, or if autoregression in y_t is near one.

In Figure 9 I show the same MSE results as in Figure 8, but increase the error variance

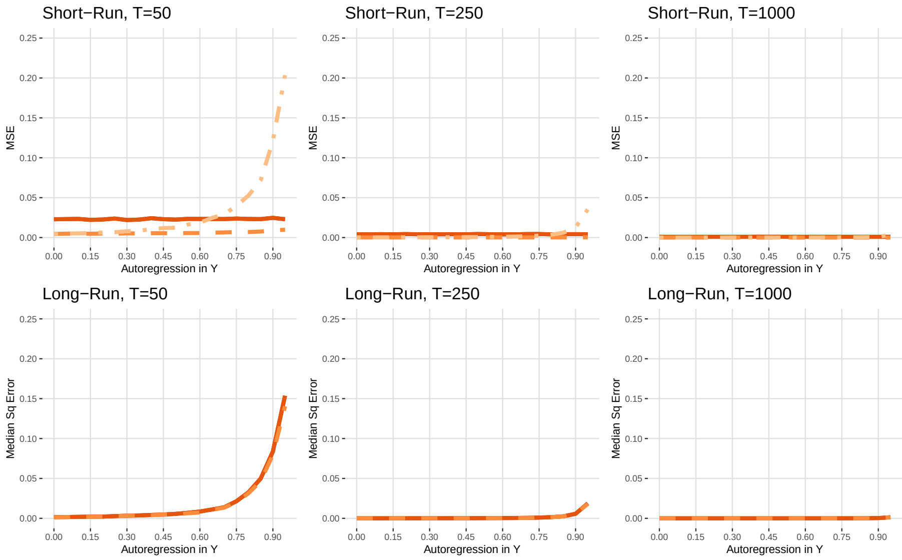

Figure 8: Scenario II: Mean (short-run) and median (long-run) square error for $Y_t \sim I(0)$, $X_t \sim I(1)$ and $\sigma_\varepsilon^2 = 1$

Note: Static (dot-dash), LDV (dash), ARDL/ECM (solid). Long-run effects do not exist for static model.

of y_t to $\sigma_\varepsilon^2 = 5$. Results are similar to those with smaller error variance, although MSE rates are approximately four times larger.

Summary: When $Y_t \sim I(0)$ and $X_t \sim I(1)$, short-run rejection rates appear to only be an issue for the static model, and do not decline as T increases (as they do in the ARDL/ECM and LDV). In the long-run, rejection rates are high for both the ARDL/ECM and LDV model but quickly approach convention as T grows or as autoregression in y_t declines. MSE in the short-run is only an issue for the static model in short T and large autoregression in y_t. In the long-run, MSE is high in both the ARDL/ECM and LDV specifications in short T and large autoregression in y_t.

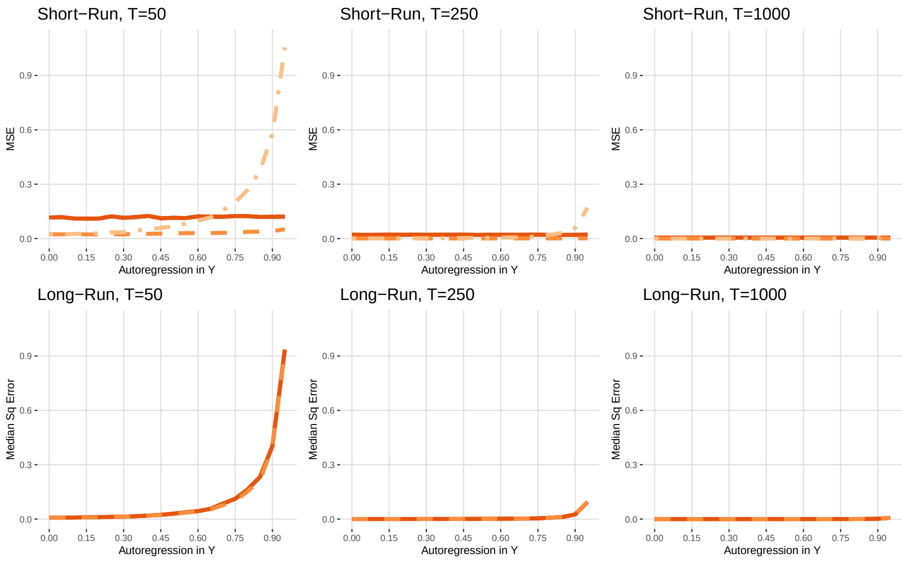

Figure 9: Scenario II: Mean (short-run) and median (long-run) square error for $Y_t \sim I(0)$, $X_t \sim I(1)$ and $\sigma^2_\varepsilon = 5$

Note: Static (dot-dash), LDV (dash), ARDL/ECM (solid). Long-run effects do not exist for static model.

1.3 Scenario III: $Y_t \sim I(1)$, $X_t \sim I(0)$, and unrelated

In Scenario III, the dependent variable now contains a unit root, while the independent variable is stationary, yet possibly autoregressive. In Figure 10, I show rejection rates under increased error variance in y_t ($\sigma_\varepsilon^2 = 5$). Results are nearly identical to those in the main paper when $\sigma_\varepsilon^2 = 1$; in the short-run, rejection rates are far above convention in the static model as autoregression in x_t increases. In the long-run, rejection rates tend to be below convention, only rising above 0.05 when $T = 50$ and autoregression in x_t is above 0.75; this applies to both the ARDL/ECM and LDV specifications.

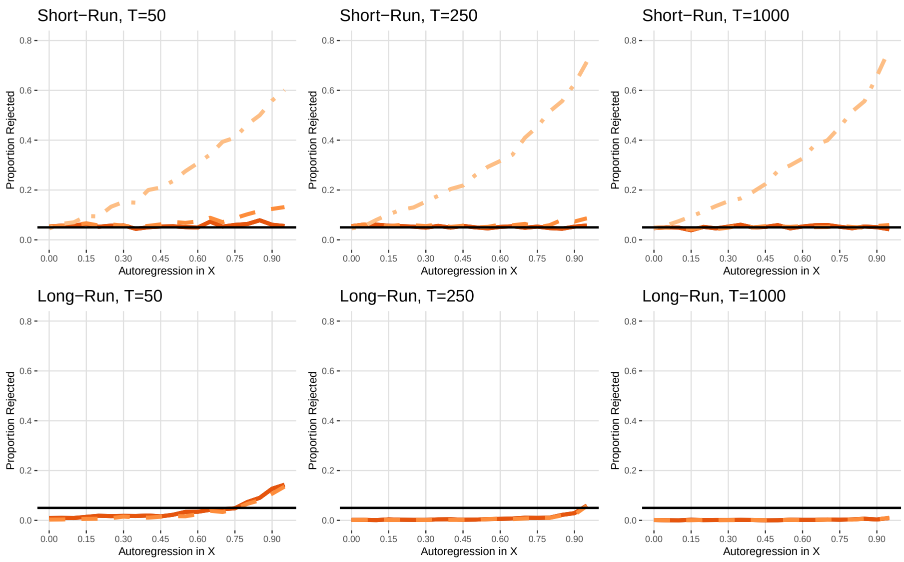

Figure 10: Scenario III: Rejection rates for $Y_t \sim I(1)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

Note: Static (dot-dash), LDV (dash), ARDL/ECM (solid). Long-run effects do not exist for static model.

In Figure 11 I show mean square error estimates for the short-run (top row) and median square error for the long-run (bottom row). As discussed briefly in the main paper, these results—when the dependent variable is I(1) but the independent variable is I(0)—are quite different from those when the order of integration is switches (i.e., Scenario II, where $Y_t \sim I(0)$, $X_t \sim I(1)$). In the short-run, MSE is quite low for the ARDL/ECM and LDV models. MSE is relatively larger in the static model, and gets worse as autoregression in x_t increases. In contrast, we see almost the opposite effect in the long-run. First, MSE is much larger in the long-run than in the short run, by several orders of magnitude. Second, MSE gets *smaller* across all models as autoregression in x_t increases. This finding adds nuance to previous studies such as De Boef and Granato (1997), since it suggests that estimates are less likely to be spurious as x_t becomes a near-unit-root (i.e., they are driven towards zero). Taken together with the rejection rates (e.g., Figure 10), the findings here suggest that for the long-run estimates ARDL/ECM and LDV models are far from zero, especially when autoregression in x_t is small, but that the confidence intervals nearly always overlap zero.

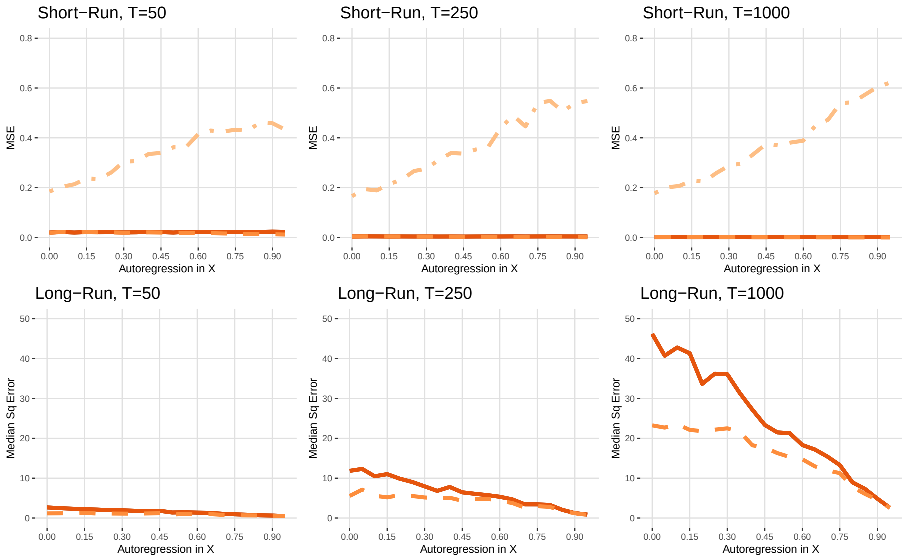

Figure 11: Scenario III: Mean (short-run) and Median (long-run) square error for $Y_t \sim I(1)$, $X_t \sim I(0)$ and $\sigma^2_\varepsilon = 1$

Note: Static (dot-dash), LDV (dash), ARDL/ECM (solid). Long-run effects do not exist for static model.

Figure 12 shows MSE results when the error variance in y_t is increased to $\sigma^2_\varepsilon = 5$. Results are quite similar to those in Figure 11, although MSE appears to be between four and five times larger in Figure 12.

Summary: When $Y_t \sim I(1)$ and $X_t \sim I(0)$ and are unrelated, in the short-run, rejection rates are only a concern for the static model, and do not improve in T. For the long-run estimates, ARDL/ECM and LDV models are far from zero, especially when autoregression in x_t is small, but that the confidence intervals nearly always overlap zero, as evidenced by relatively small rejection rates but very large MSE values. Thus, when $Y_t \sim I(1)$ and $X_t \sim I(0)$ and not related, large long-run effects that are not statistically significant seem common. In contrast (see Scenario II), if $Y_t \sim I(0)$ and $X_t \sim I(1)$ and are unrelated, rejection rates of the long-run effect are higher, although they tend not to be far away

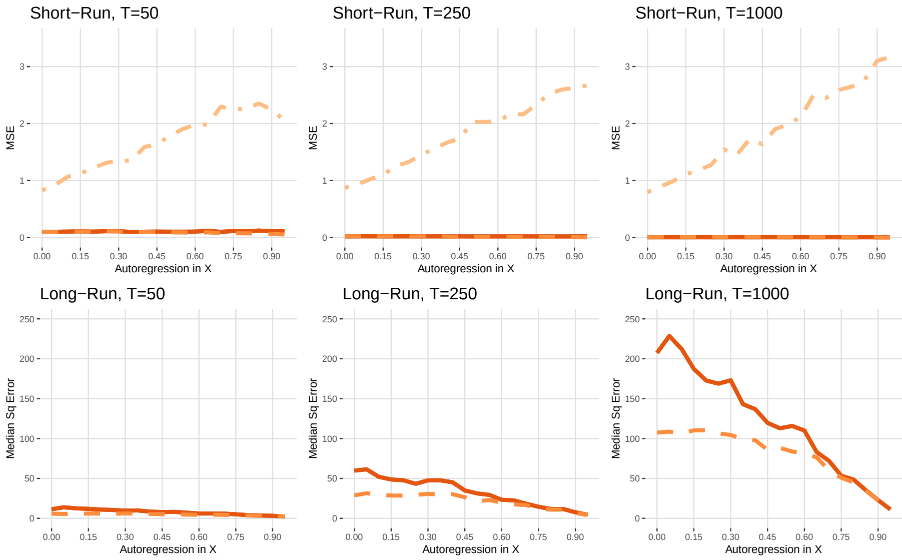

Figure 12: Scenario III: Mean (short-run) and Median (long-run) square error for $Y_t \sim I(1)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

Note: Static (dot-dash), LDV (dash), ARDL/ECM (solid). Long-run effects do not exist for static model.

from zero. This is likely because in the former, the coefficient on the lag of y_t will be significant and near one; this will produce very large long-run effects (i.e., dividing by a small, near-zero value: $1-\hat{\alpha} \to 0$ if $\hat{\alpha}$ is near one). Yet the coefficients on x_t tend to not be statistically significant, which result in large standard errors surrounding a large (far from zero) long-run effect estimate. In contrast, if $Y_t \sim I(0)$ and $X_t \sim I(1)$ (Scenario II), both the coefficient on the lag of y_t (since the series is often autoregressive) and sometimes the coefficients on x_t appear to often be statistically significant, meaning that the long-run effect is typically small, but sometimes statistically significant.

1.4 Scenario IV: $Y_t \sim I(1)$, $X_t \sim I(1)$, and unrelated

In Scenario IV, I regress two unrelated series containing a unit root on one another. In Figure 13, I increase the error variance of y_t to $\sigma_\varepsilon^2 = 5$. Results are similar to those in the main paper where the error variance was only $\sigma_\varepsilon^2 = 1$, although while in the main paper the ARDL, ECM, and LDV models appeared to perform slightly worse in correct coverage of the long-run effect in increasing T (likely because of smaller standard errors achieving levels of statistical significance more often), here there is less clear of a relationship as T grows. Regardless, rejection rates are well above convention for the LDV and static model for the short-run effects, and above convention (evidence of spurious findings 20 percent of the time) for all three models in the long-run.

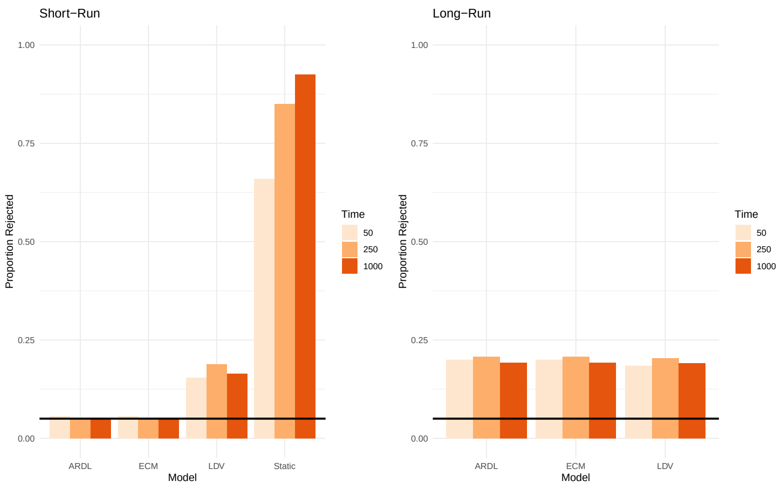

Figure 13: Scenario IV: Rejection rates for $Y_t \sim I(1)$, $X_t \sim I(1)$ and $\sigma_\varepsilon^2 = 5$

In Figure 14, I show mean square error estimates for the short-run effects when both series contain a unit root. The left plot in the figure shows MSE when $\sigma_\varepsilon^2 = 1$, and $\sigma_\varepsilon^2 = 5$ in the right plot. MSE is clearly highest in the static model, and lowest (only slightly) for the LDV model. When error variance in y_t increases to five, MSE gets larger by over a factor of four. Moreover, while MSE declines in the ARDL, ECM and LDV models as T increases, it appears to either stay the same or possibly grow for the static model.

In Figure 15, I show median square error results for the three models that can recover long-run effects, when both variables are I(1). MSE appears to be similar across the model specifications—note that as mentioned in the main paper, ARDL and ECM results are

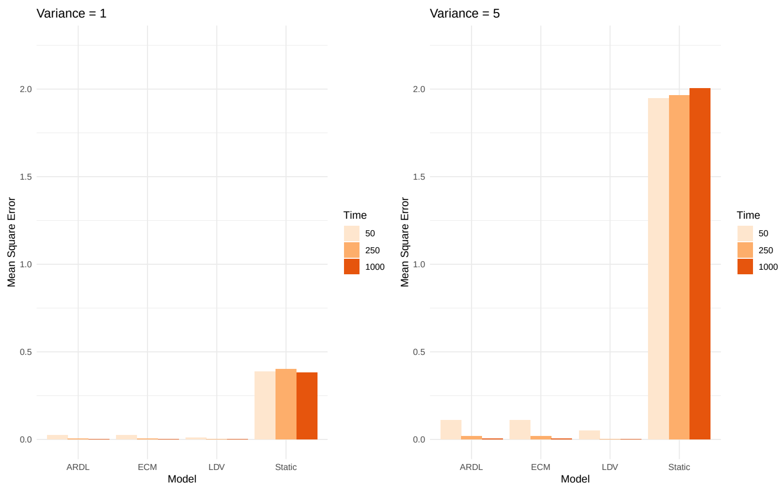

Figure 14: Scenario IV: Mean squared error for short-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

identical—and perhaps increasing slightly in T when $\sigma^2_\varepsilon = 1$, and decreasing when $\sigma^2_\varepsilon = 5$.

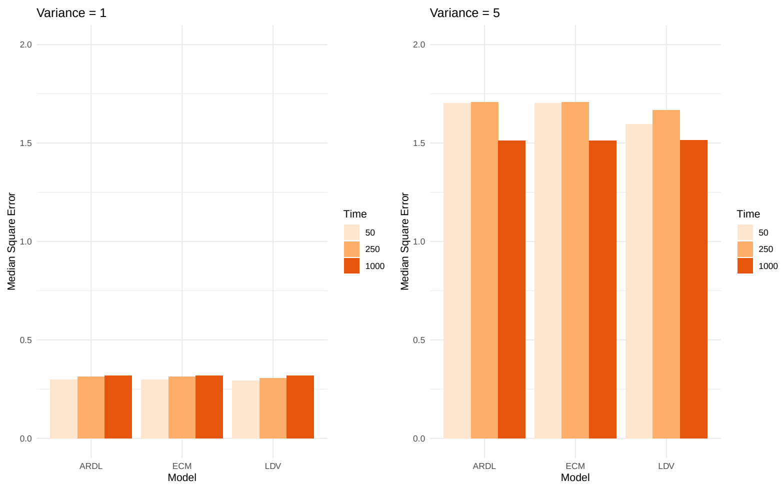

Figure 15: Scenario IV: Median squared error for long-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

Summary: When two I(1) series are spuriously related and regressed one another, short-run estimates appear to be correct for the ARDL/ECM specification (and close for the LDV), although long-run effects are not correct, and get worse as the error variance in y_t increases. Short-run rejection rates and MSE are extremely high for the static model, as others have found (Granger and Newbold 1974; De Boef and Granato 1997).

1.5 Scenario V: $Y_t \sim I(0)$, $X_t \sim I(0)$, and related

In this scenario, both variables are stationary and related. The quantities that are varied in this scenario are the coefficient on the lag of x_t (either -1 or 1), the level of autoregression in the dependent variable ($\alpha = 0.2, 0.8$) and the length of the series. The short-run effect is fixed at $\beta_1 = 2$, while the long-run effect varies based on the value of α and β_2.[1]

In Figure 16, I show rejection rates of the short- and long-run effects; this is the same as the main paper, except here I increase the error variance of the dependent variable to $\sigma_\varepsilon^2 = 5$. Ideally we should expect to reject the true values of the effects around five percent of the time For the short-run effects (left plots in Figure 16), rejection rates are around 0.05 for all model specifications; only when the dependent variable is highly autoregressive ($\alpha = 0.8$) and the lagged effect of x_t is one does the static model rise above conventional levels, and only to about six or seven percent rejection.

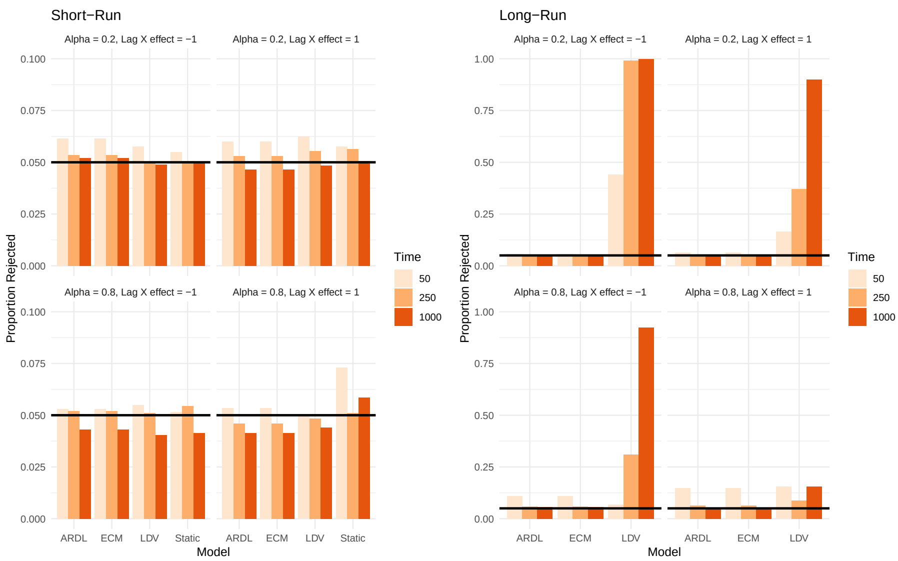

Figure 16: Scenario V: Rejection rates of the effects, $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma_\varepsilon^2 = 5$

In the long-run (right plots in Figure 16), the ARDL/ECM is around convention when the level of autoregression is low ($\alpha = 0.2$), no matter whether the effect of the lag of x_t is positive or negative. In contrast, the LDV model has rejection rates approaching 100

[1] For instance, the long-run effect when $\alpha = 0.2$ and $\beta_2 = 1$ would be $LRM = \frac{2+1}{(1-0.2)} = 3.75$, but 15 if $\alpha = 0.8$ and $\beta_2 = 1$.

percent of the time, and this appears to get worse in T. Moreover, rejection rates for the LDV are higher when the coefficient on the lag of x_t is of oppositely sign than that of x_t (-1 vs 2). A similar high rejection rate for the LDV also occurs when y_t is highly autoregressive and the lag of x_t is negative. For the ARDL/ECM, rejection rates are slightly above convention when y_t is highly autoregressive, yet this quickly declines in T. And when the lag of x_t is in the same direction as the short-run effect and y_t is highly autoregressive (bottom-right), the LDV performs roughly similar to the ARDL/ECM.

In Figures 17 and 18, I show power, or how often the null hypothesis of no effect is rejected. Since the series are related, higher power is more preferred. Across all models, levels of autoregression in y_t, and effect of the lag of x_t, power appears to be well above 0.95, with the exception of the LDV (when $\alpha = 0.8$ and $\beta_2 = 1$), and the long-run effects for the ARDL/ECM model in short series and high error variance in y_t (when $\alpha = 0.8$ and $\beta_2 = 1$). Given these findings (i.e., depending on the coefficients chosen in the Monte Carlo, we could probably always have—or never have—Type II error), I chose to present rejection rates of the actual calculated effects in the main paper.

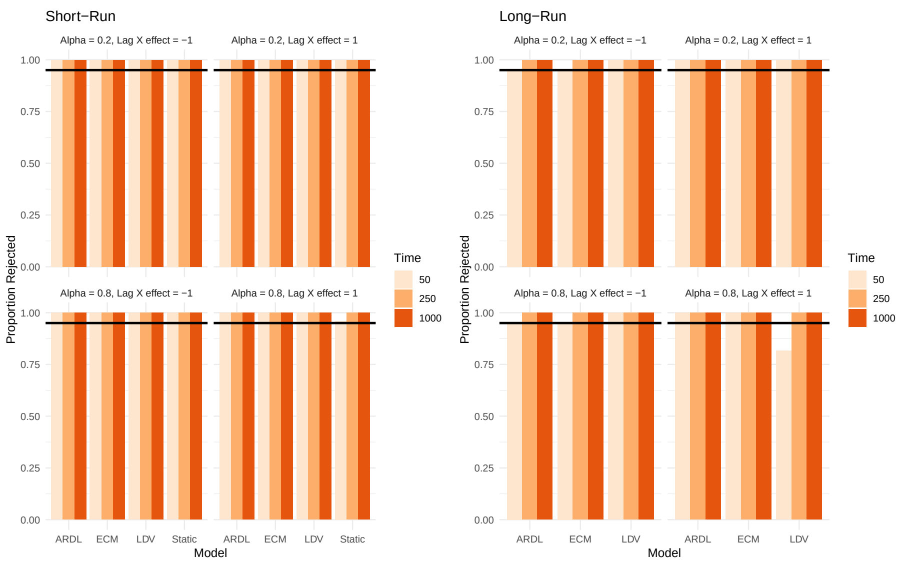

Figure 17: Scenario V: Power (rejection rates of zero) $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma^2_\varepsilon = 1$

In Figure 19, I show estimates of the mean square error of the short-run effects for the two related series. When the error variance of y_t is low (left plots in Figure 19), MSE is very low for all models except for the static model under high autoregression and

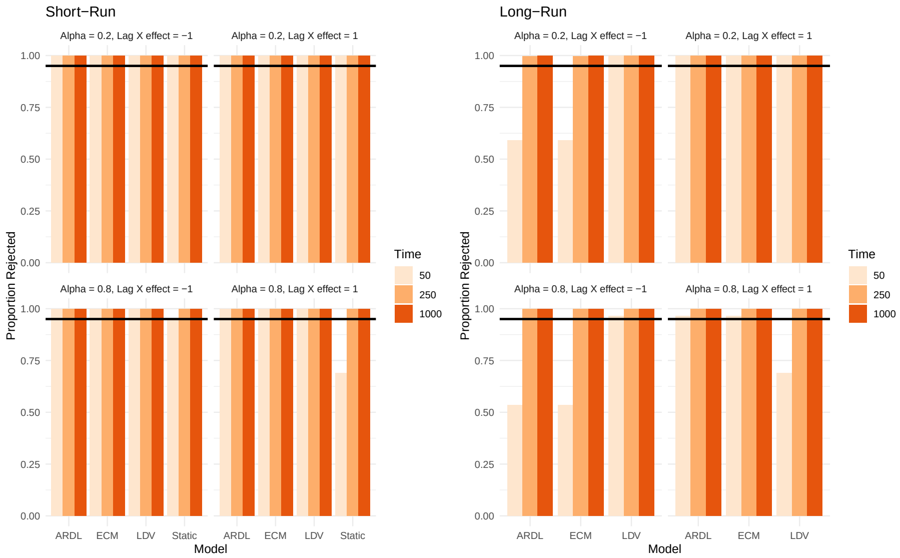

Figure 18: Scenario V: Power (rejection rates of zero) $Y_t \sim I(0)$, $X_t \sim I(0)$ and $\sigma^2_\varepsilon = 5$

a positive lag effect of x_t (when $\alpha = 0.8$ and $\beta_2 = 1$). Across all models, MSE declines quite quickly as T increases; even in the static specification MSE becomes negligible (and very similar to the other models) once $T = 1000$. And, in general, as the error variance increases the MSE also increases, as shown by the right plots in Figure 19.

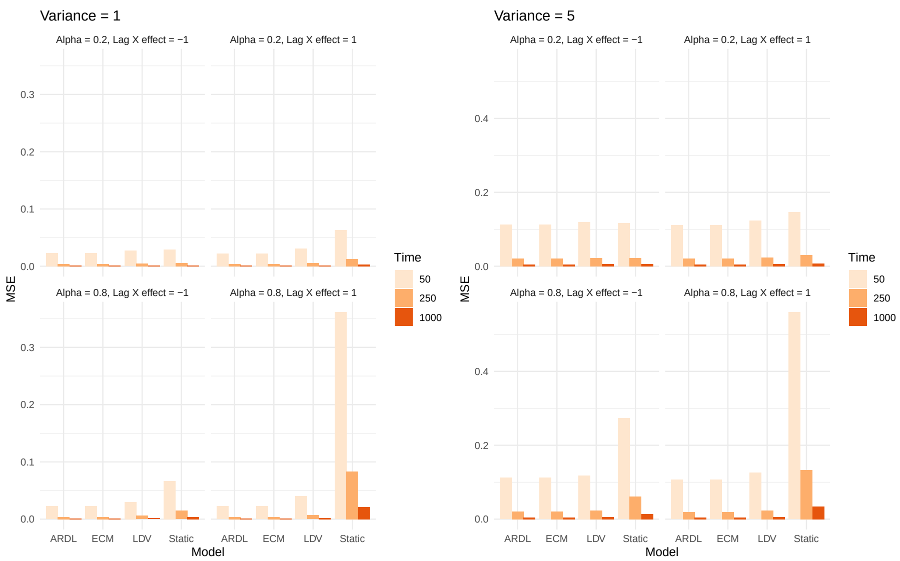

Figure 19: Scenario V: Mean square error of the short-run effect, $Y_t \sim I(0)$, $X_t \sim I(0)$

In Figure 20 I show median square error for the long-run effects for the stationary yet related scenario. When y_t is weakly autoregressive, MSE is quite low across all specifications, no matter the direction of the effect of x_{t-1} or the amount of error variance in y_t (MSE is slightly larger for the LDV specification than the ARDL/ECM). When y_t is highly autoregressive ($\alpha = 0.8$), the results change substantially. Under this condition, the ARDL/ECM models still have low MSE as long as T is large, although MSE is much higher in small T when the lagged effect of x_t is in the same direction as the short-run effect (e.g., the "Alpha = 0.8, Lag X effect = 1" scenario shown in Figure 20). When the error variance in y_t is large, while a similar pattern arises for the ARDL/ECM models, for the LDV model, things look somewhat different; when the lagged effect of x_t is negative, the model performs well, although not quite as well as the ARDL/ECM. When the lagged effect of x_t is positive, MSE is much higher, although this appears to decline more quickly in T when the error variance of y_t is large.

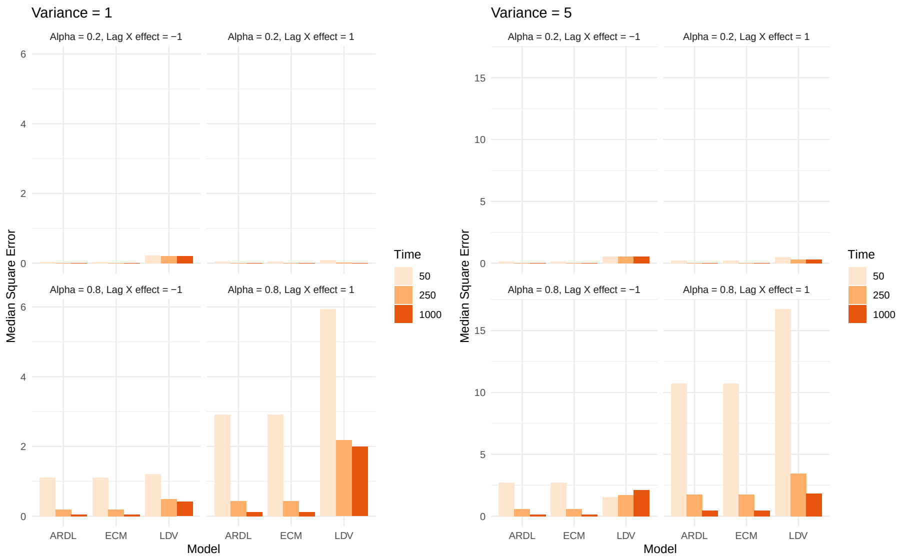

Figure 20: Scenario V: Median square error of the long-run effect, $Y_t \sim I(0)$, $X_t \sim I(0)$

In the following figures in this section, I show box-plots of the estimates of the short-run and long-run effects. These are an effective way to better understand the other quantities of interest presented throughout the paper, such as mean squared error and rejection rates, since they show the actual distribution of the estimates. They also aid in determining if, on average, estimates are consistently falling below or above a desired quantity. Each quadrant of plots shows a different model, each sub-plot in a quadrant varies the level of α, the two different coefficients on the lag of x_t appear on the horizontal axis of these plots, and the actual box-plots shown differ in T.

Figure 21, which shows the distribution of short-run effects, appears to be estimated without bias for all models except the static model. When the lagged dependent variable is weakly autoregressive ($\alpha = 0.2$), estimates appear to be attenuated when the coefficient on the lag of x_t is positive. This appears to get larger as α increases (as shown by the right plots for the static model), where now attenuation occurs both when $\beta_2 = -1$ and $\beta_2 = -1$. Note that this only appears to occur when T is small; short-run effects are recovered fairly well for all models in large T.

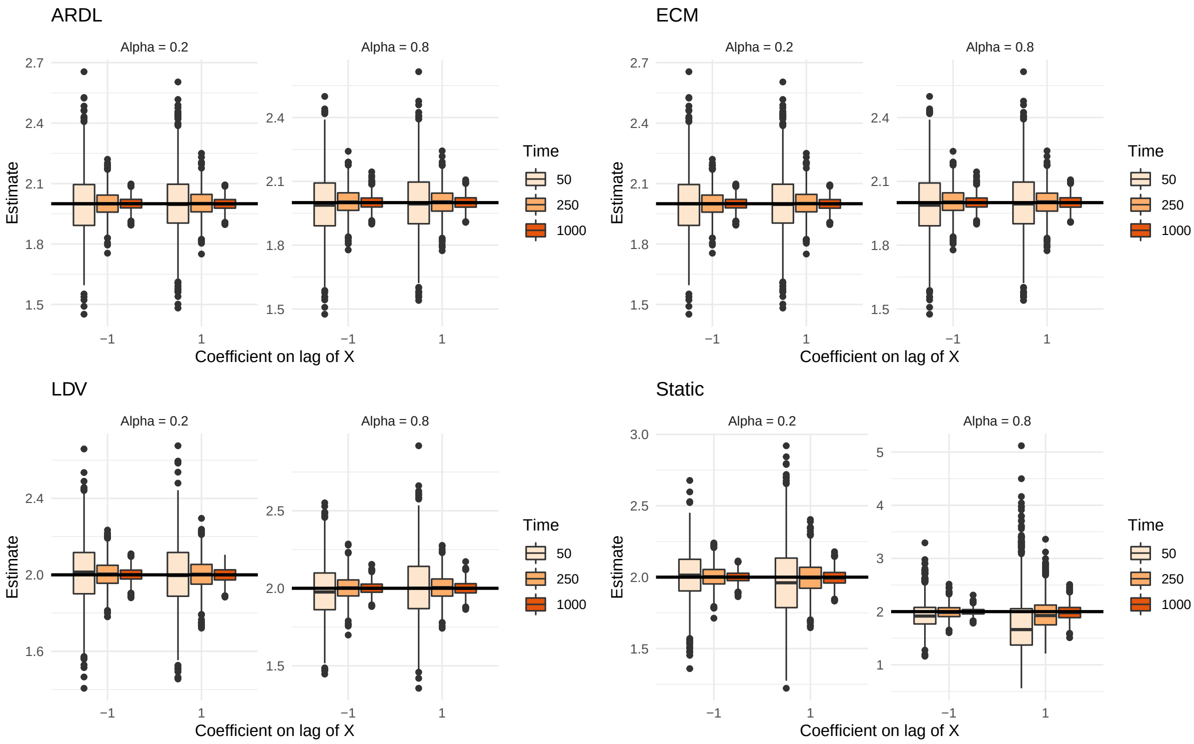

Figure 21: Scenario VI: Box-plots of the short-run effect, $Y_t \sim I(0)$, $X_t \sim I(0)$

Note: $\sigma_\varepsilon^2 = 1$ for all simulations shown.

In Figure 22 I show box-plots of the estimates of the long-run effect. Note that since each of the true long-run effects differs based on the value of α and β_2, for comparability I center each of the long-run effect estimates such that a value of zero means that the estimate is perfectly centered on the true long-run effect. The ARDL/ECM model appears unbiased and consistent as T increases, although long-run estimates are attenuated slighly in small T when $\alpha = 0.8$, especially when the coefficient on the lag of x_t is of opposite sign as the coefficient on x_t (i.e., $\beta_2 = -1$).

In contrast to the ARDL/ECM model, the LDV results in Figure 22 suggest that

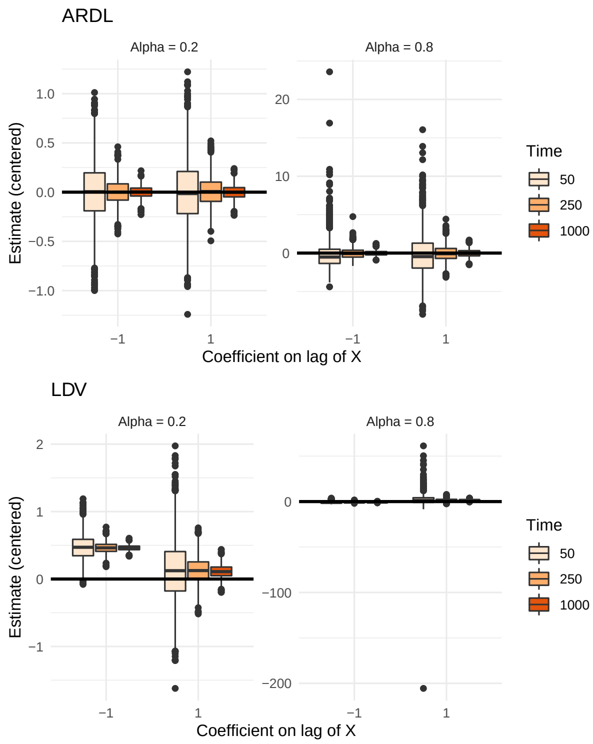

Figure 22: Scenario VI: Box-plots of the long-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

Note: Long-run effects do not exist for static model. Estimates are centered around zero to enable comparison across different long-run effects given the combination of α and β_2. $\sigma_\varepsilon^2 = 1$ for all simulations shown.

when the dependent variable is only slightly autoregressive ($\alpha = 0.2$) and when $\beta_2 = -1$, estimates are nearly always biased upwards from the correct long-run effect. Moreover, estimates remain inconsistent as T increases. There is less bias when $\beta_2 = 1$, although a similar issue exists. A similar result appears for the LDV when $\alpha = 0.8$, although the results are harder to see given the extremely large single estimate of a very low long-run effect.

Summary: When two series are stationary and related, the ARDL/ECM model appears to be the best choice, although coverage of the true effects and power are affected in small T. Long-run effects suffer from high median-square when α is large, and much less so when it is small. Short-run effects can be accurately obtained by all models, although the static model is the worst performer. The LDV model has extremely poor coverage of the long-run effects compared to the ARDL/ECM models, especially when α is low or the coefficient on the lagged dependent variable is negative. Long-run effects are nearly always over-estimated (away from zero) for the LDV model (no matter T), while they can be attenuated towards zero in the ARDL/ECM when T is small.

1.6 Scenario VI: $Y_t \sim I(1)$, $X_t \sim I(1)$, and cointegrating

In this scenario, I present additional results from two I(1) series that are in a cointegrating relationship. in Figure 23, I show rejection rates of the short- and long-run effects, similar to those in the main paper, but now increasing the error variance in y_t to $\sigma_\varepsilon^2 = 5$. The results are nearly identical for the short-run effect; constructed 95 percent confidence intervals for the short-run effect for the LDV and static models fail to encompass the true effect of $\beta_1 = 2$ nearly 100 percent of the time. Rejection rates are slightly higher for the long-run effects under increased error variance; all models suffer from higher rejection rates of the true long-run effect when the error correction mechanism is highly persistent (i.e., $\alpha = -0.2$), especially in small T. The LDV model has rejection rates higher than convention for all scenarios except for when $\alpha = -0.8$, T is large, and the coefficient on the lag of x_t is -1.

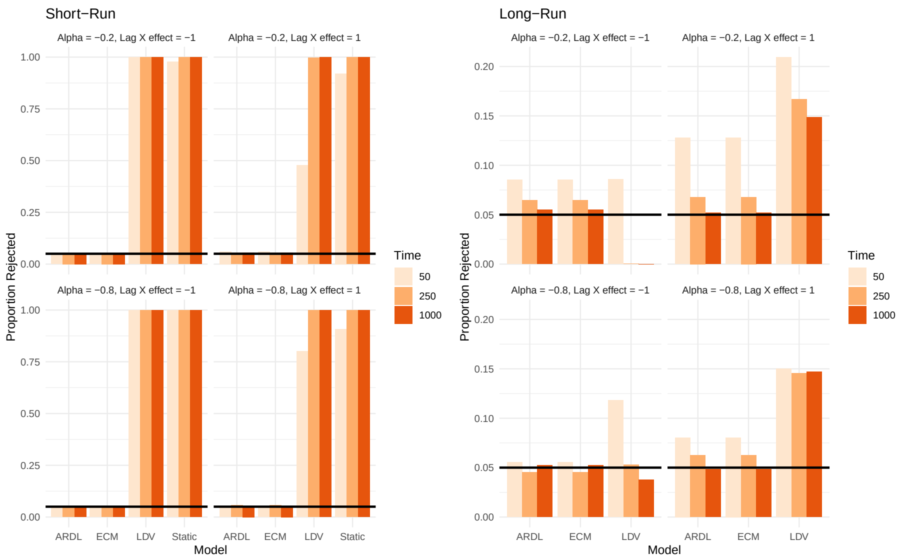

Figure 23: Scenario VI: Rejection rates of the effects, $Y_t \sim I(1)$, $X_t \sim I(1)$ and $\sigma_\varepsilon^2 = 5$

In Figures 24 and 25 I show the power (the proportion of times the estimated effect's constructed 95 percent confidence intervals do not include zero) across all scenarios when $\sigma_\varepsilon^2 = 1$ and $\sigma_\varepsilon^2 = 5$ for each of the figures, respectively. Similar to Scenario V, the way in which the Monte Carlo was constructed meant that power was high for nearly every model in most circumstances, making rejection rates of the true effects (e.g., Figure 23) substantively more interesting. As is clear from Figures 24 and 25, power is very high except for the LDV and static model in small-T cases in which the error correction mechanism is not persistent ($\alpha = -0.8$) and the coefficient on the lag of x_t is $\beta_2 = -1$.

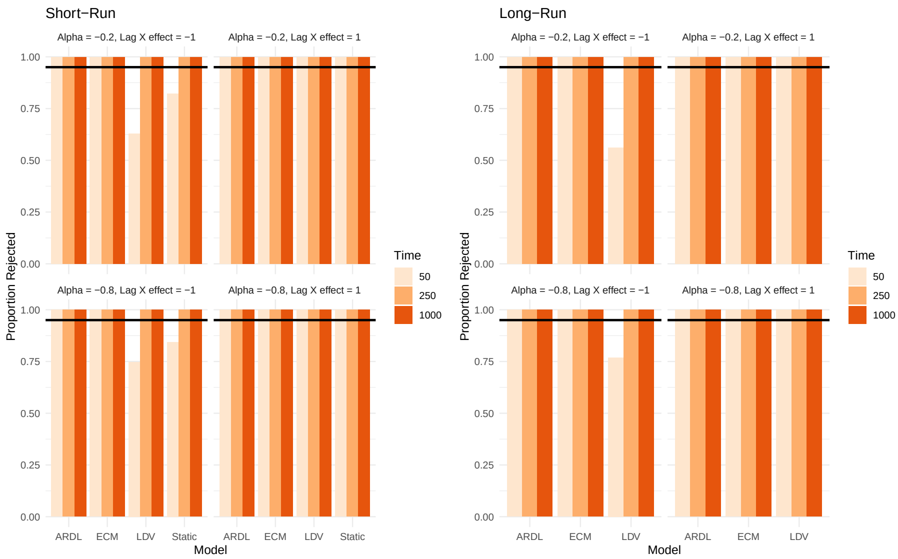

Figure 24: Scenario VI: Power (rejection rates of zero) $Y_t \sim I(1)$, $X_t \sim I(1)$ and $\sigma^2_\varepsilon = 1$

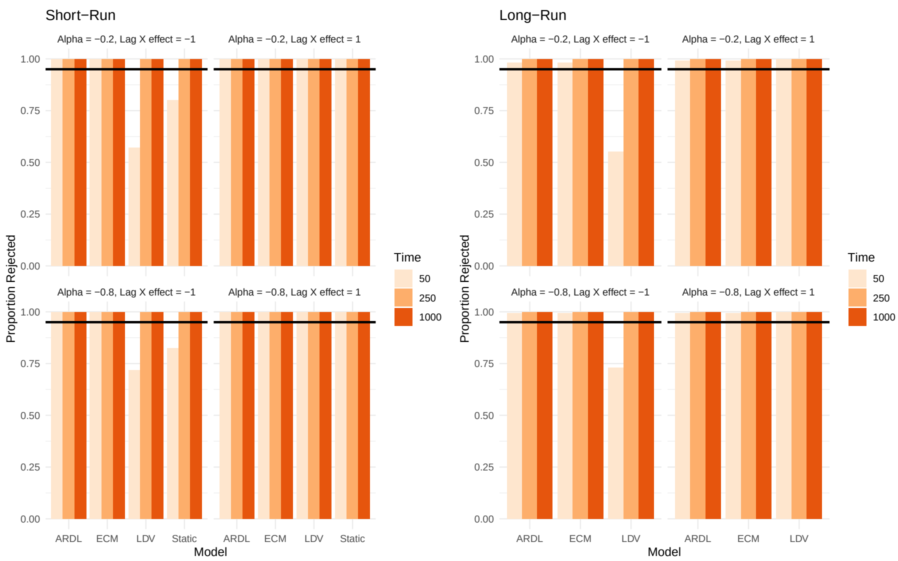

Figure 25: Scenario VI: Power (rejection rates of zero) $Y_t \sim I(1)$, $X_t \sim I(1)$ and $\sigma^2_\varepsilon = 5$

In Figure 26 I show mean square error results for the short-run effect under the cointegrating scenario. MSE is very low for the ARDL/ECM under all scenarios. Short-run MSE appears to be worst overall for the static model, and then the LDV model. MSE for these two models is worst when the error correction mechanism is highly persistent ($\alpha = -0.2$) and the coefficient on the lag of x_t is in the opposite direction as that on x_t (i.e., $\beta_2 = -1$ while $\beta_1 = 2$). Interestingly, short-run MSE appears to be nearly identical, no matter the error variance in y_t.

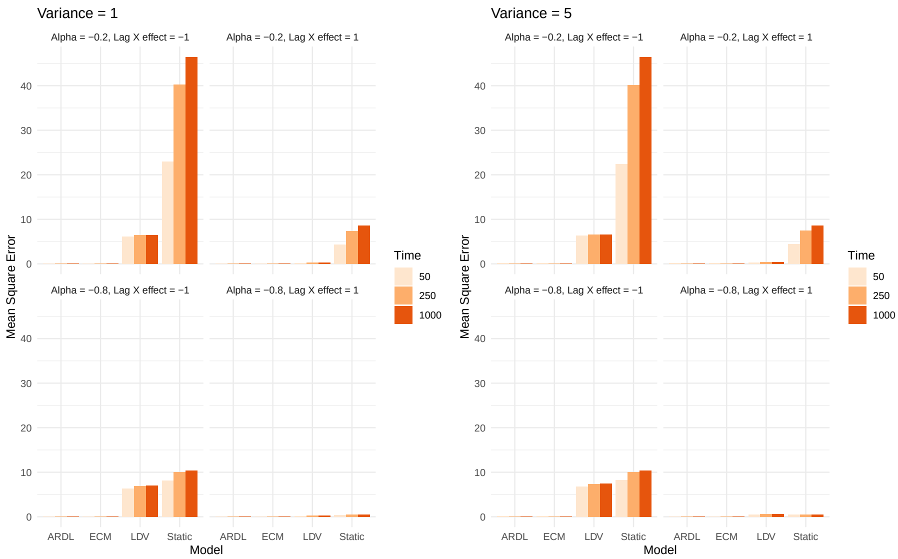

Figure 26: Scenario VI: Mean square error of the short-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

Note: Figure shows median square error of the short-run effect for the models across differing levels of error variance, σ^2_ε, persistence in the error-correction rate, α, coefficient on x_{t-1}, β_2, and T.

In Figure 27, I show the median square error results for the long-run effect under cointegration. Looking at the $\sigma^2_\varepsilon = 1$ results (left plots in Figure 27), it is clear that median square error appears to only be an issue in any of the models when $T = 50$. Moreover, this is substantially worse for the LDV specification, but only when the error-correction mechanism is highly persistent and the coefficient on the lag of x_t is in the opposite direction as that on x_t (i.e., $\beta_2 = -1$ while $\beta_1 = 2$); in this instance, MSE is about 10 times larger for the LDV than for the ARDL/ECM when $T = 50$. MSE increases across all models when the error variance around y_t increases, in contrast to the short-run effects when there was no discernible difference.

In Figure 28 I show box-plots of the estimates of the short-run effect. As described in the previous section, these show the actual distribution of the estimates, which can aid in determining if, on average, estimates are consistently falling below or above a desired quantity. For the ARDL/ECM specifications, short-run effects appear centered on the true value of two and estimates appear consistent as T increases. For the LDV and static models, these box-plots go a long way in explaining why MSE was so high under cointegration, as we found in Figure 26. When the coefficient on the lag of x_t is one, the

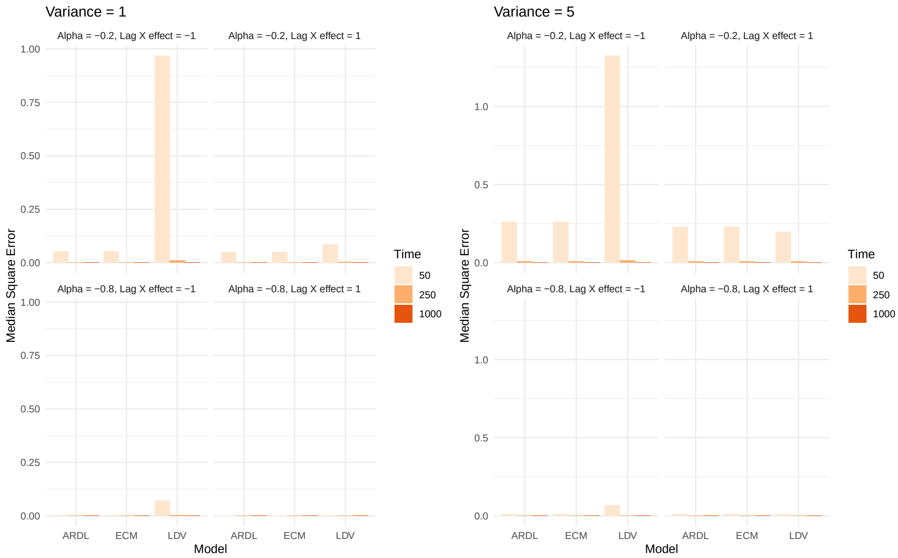

Figure 27: Scenario VI: Median square error of the long-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

Note: Figure shows median square error of the long-run effect for the models across differing levels of error variance, σ^2_ε, persistence in the error-correction rate, α, coefficient on x_{t-1}, β_2, and T. Long-run effects do not exist for static model.

LDV model produces attenuated short-run estimates of about 1.5, no matter the level of α. This makes sense; if the short-run effect of the DGP is $\beta_1 = 2$ and the coefficient on the lag of x_t is $\beta_2 = 1$, the LDV—which is estimating a single coefficient for x_t—will produce some compromise between these two effects. This becomes even more clear when $\beta_2 = -1$. In these cases, the esimate of the short-run effect for the LDV in Figure 28 is now negative, far away from the true value of two. Of course, the LDV *is* including a lagged dependent variable, so the error correction mechanism does not appear to affect short-run effects, only the relationship between the coefficients on x_t and x_{t-1}.

For the static model shown in Figure 28, which, unlike the LDV, does not include a lagged dependent variable, the results of the short-run effect are even worse. When $\beta_2 = -1$, short-run estimates are negative; this gets worse when the error correction mechanism is highly persistent ($\alpha = -0.2$). When $\beta_2 = 1$ (i.e., is in the same direction as the short-run effect), estimates in the static model are closer to the true value of two, but are underestimated when the error correction mechanism is not persistent (i.e., $\alpha = -0.8$), and overestimated when it is. Across all scenarios for both the static and LDV specifications, estimates appear to get worse (i.e., converge on the incorrect short-run

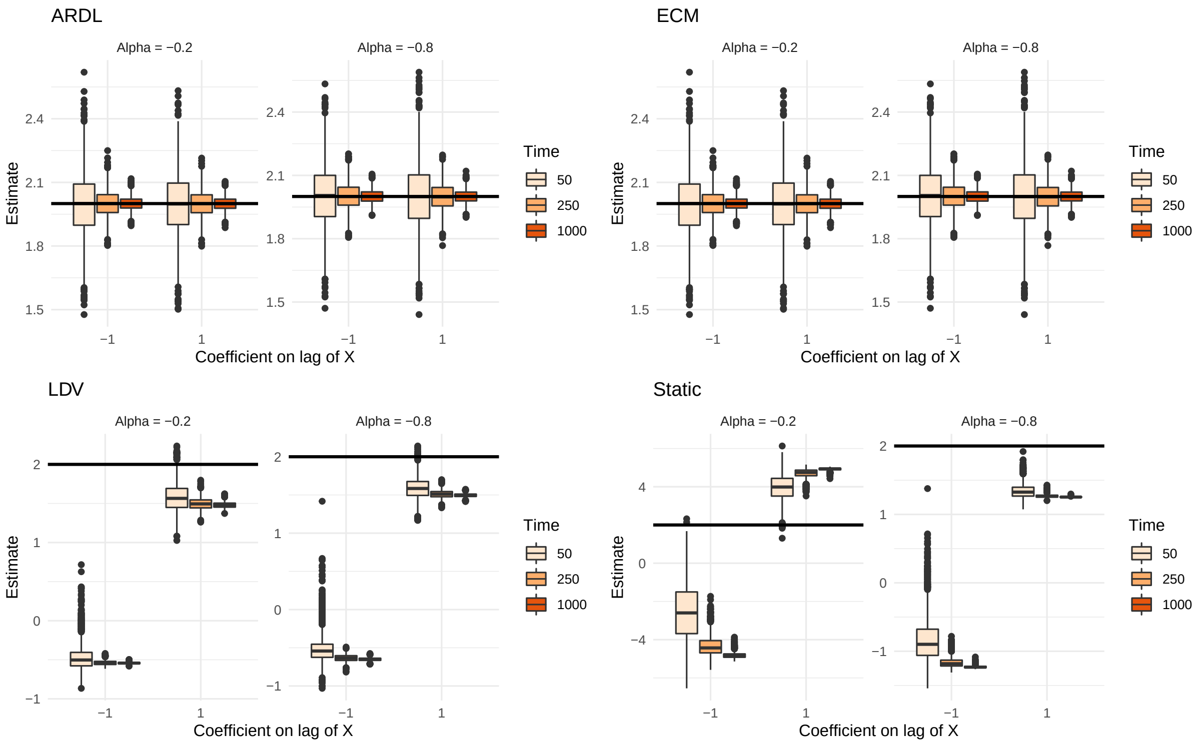

Figure 28: Scenario VI: Box-plots of the short-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

Note: $\sigma_\varepsilon^2 = 1$ for all simulations shown.

value) as T increases.

In Figure 29 I show box-plots of the estimates of the long-run effect. Since each of the true long-run effects differs based on the value of α and β_2, I center each of the long-run effect estimates such that a value of zero means that the estimate is perfectly centered on the true long-run effect, to ensure comparability. Once again, the ARDL/ECM specifications have long-run estimates that appear to be centered on the true value of the long-run effects. The LDV also appears to do a fairly good job at recovering, on average, unbiased estimates of the long-run effect, especially if the coefficient on the lag of $x_t = 1$. When the error correction mechanism is not persistent (i.e., $\alpha = -0.8$), long-run estimates for the LDV are biased upwards, and overall long-run estimates have a much larger spread than for the ARDL/ECM specification.

Summary: The ARDL and ECM models are identical throughout this and all other data-generating processes. When the dependent and independent variables are cointegrating, constructed 95 percent confidence intervals for the static and LDV model fail to

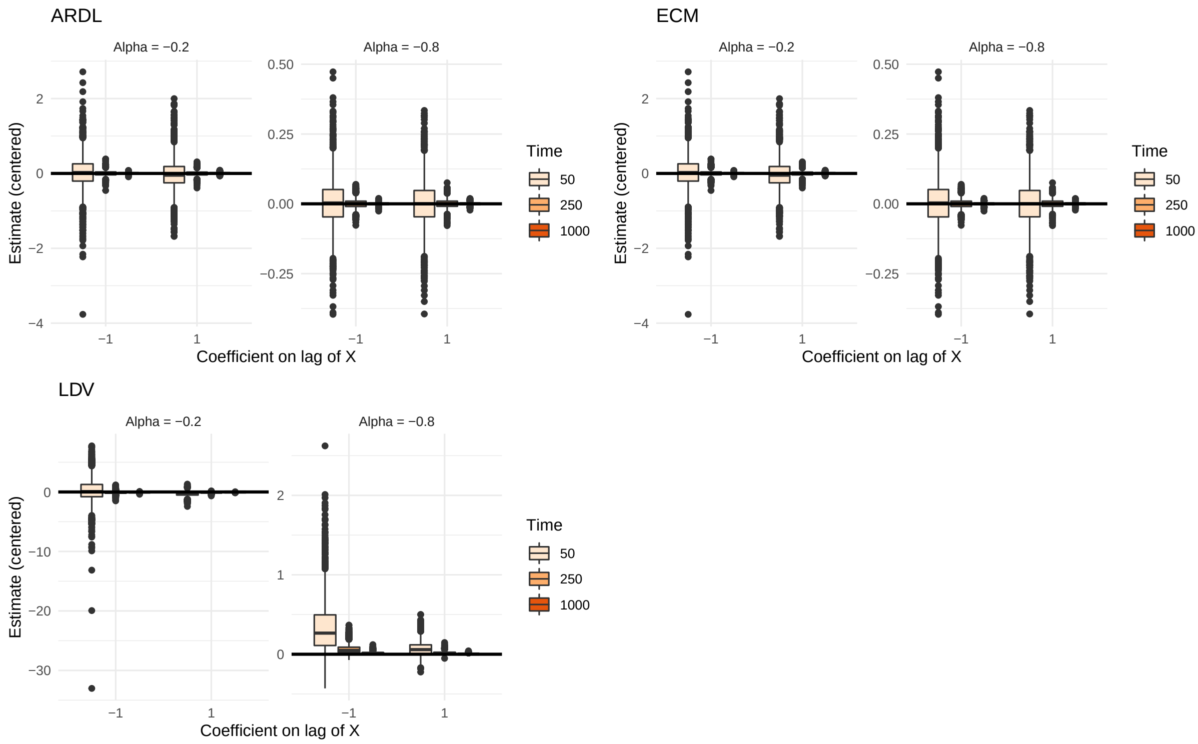

Figure 29: Scenario VI: Box-plots of the long-run effect, $Y_t \sim I(1)$, $X_t \sim I(1)$

Note: Long-run effects do not exist for static model. Estimates are centered around zero to enable comparison across different long-run effects given the combination of α and β_2. $\sigma_\varepsilon^2 = 1$ for all simulations shown.

encompass the true short-run effect nearly 100 percent of the time, nor does this improve in T. In other words, when data are cointegrating, obtaining correct short-run effects, which previous results showed was sometimes possible with the LDV and static specifications (when data were unrelated or related but stationary and autoregressive), is nearly impossible. In terms of directions of this bias, in the LDV model it is a compromise of the coefficients on x_t and x_{t-1} (since only a single coefficient for x_t is estimated), while in the static model the direction and size of the bias depends on the rate of error-correction as well. In contrast, short-run effects for the ARDL/ECM specification appear unbiased and consistent in T. Long-run effects for the LDV specification have substantial spread, but tend to be fairly close to the correct value (though still far worse than the ARDL/ECM). Long-run effects for the ARDL/ECM specification appear unbiased and consistent in T.

2 Alternative strategies for long-run effect inferences

In the discussion section in the manuscript I mentioned two possible strategies that seem tempting to use, given the Monte Carlo results showing that short-run inferences tend to always be correct (when estimating the ARDL/ECM), while the same cannot be said for long-run effects. The first is whether we should proceed to test for a long-run effect if we do not find evidence of a short-run effect. The second—which is a strategy suggested by Enns, Moehlecke and Wlezien (henceforth EMW) in this symposium—suggests not calculating a long-run effect if the coefficient on the lag of the regressor is not found to be statistically significant. While page constraints limit this exposition in the manuscript, I elaborate on why these two alternative strategies are not advisable below.

2.1 First short-run, then long-run?

Can there be a long-run effect if there is no short-run effect? To be clear, by "short-run", I mean the contemporaneous effect of x on y at time t. While the short-run effect is part of the long-run effect, there still can be a long-run effect when no short-run effect is observed. Consider a simple example where $x_t \sim N(0,1), \varepsilon_t \sim N(0,1)$, and:

$$y_t = \phi y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t = 0.8 y_{t-1} + 0 x_t + 3 x_{t-1} + \varepsilon_t \tag{1}$$

An ARDL(1,1) regression using simulated data of this process is shown in Table 1. Given the predominant interpretation of short-run effects, the non-significant coefficient on x_t would lead us to conclude that there is indeed no short-run effect. Yet clearly a long-run effect exists of $LRM = \frac{-0.02+3.12}{1-.80} = 15.44$; moreover, this effect is statistically significant. Note that we would reach the same conclusion were we to estimate an ECM as well, as shown below in Table 2. Here, we would conclude that there is no short-run effect since the coefficient on Δx_t is near-zero, although once again there is a long-run effect of $LRM = \frac{3.10}{-.20} = 15.44$.

Table 1: Simulated data with no short-run effect, ARDL(1,1)

	b (se)
y_{t-1}	0.80* (0.02)
x_t	-0.02 (0.11)
x_{t-1}	3.12* (0.11)
Constant	0.06 (0.11)

$T = 99$. Coefficients with standard errors in parentheses. $*$: $p < .05$.

Table 2: Simulated data with no short-run effect, ECM

	b (se)
y_{t-1}	-0.20* (0.02)
Δx_t	-0.02 (0.02)
x_{t-1}	3.10* (0.17)
Constant	0.06 (0.11)

$T = 99$. Coefficients with standard errors in parentheses. $*$: $p < .05$.

Thinking more broadly, the short-run effect is only *one* parameter in the LRM calculation; given L lags of x_{t-l} (and assuming an ARDL form like Equation 1):[2]

$$LRM = \frac{\sum_{l=0}^{L} \beta_l}{1-\phi} \tag{2}$$

Clearly if $\beta_0 = 0$, that does not tell us much about the values of the other parameters. Thus, while *some* β_l have to have an effect in order for there to be an LRM, it does not necessarily need to be β_0, the coefficient at time t (i.e., the short-run effect). Indeed, many models do not even estimate contemporaneous/short-run effects. Dead-start models (which was used in the example above) have regressors appearing in the model only after some lag (c.f., De Boef and Keele 2008)—in effect, restricting the short-run effect to zero—yet of course they can have long run effects. Or take cointegrating ECM models, where is not unheard of to include only lagged changes of regressors; for instance, Engle and Granger (1987, p. 272) include lagged changes from $t-1$ to $t-4$, precluding the possibility of any short-run effect appearing at time t. Indeed, in standard representations of cointegrating systems, first-differences of variables can *only* appear at time t if we assume weak exogeneity (Enders 2010).

Last, results in seminal time series texts and articles often find evidence of a statistically significant long-run effect but no short-run effect. In their error-correction model, Pesaran, Shin and Smith (2001, Table III, p. 314) find that only two of the four regressors in their model have statistically significant first differences appearing contemporaneously, even though they find evidence for cointegration. Pickup (2014, p. 192-193) estimates an ECM on all-stationary data and concludes that "there does not appear to be a significant short-run effect for any of the three economic variables". He then finds evidence that one of these variables does have a significant long-run effect.

[2]This becomes even more clear in the ECM, where the equation for the long-run effect does not even include short-run changes: $LRM = \frac{\beta_1}{-\phi}$ (where β_1 is the coefficient on the lag of x_t).

2.2 First x_{t-1}, then long-run?

The other proposed strategy involves first testing whether the coefficient on the lag of x_{t-1} is statistically significant (Enns, Moehlecke and Wlezien, this symposium). If so, proceed to calculate the long-run effect. If not, then conclude that there is no long-run effect. What if we had pursued this strategy? Below I show the same setup for the Monte Carlo's described in Section 2.1 above, but now show the rejection rates of the long-run effect equaling zero in Scenarios I-IV. In Scenarios V and VI, which estimate true relationships, I see how the EMW strategy compares to always calculating the LRM.

2.2.1 Scenario I: $Y_t \sim I(0)$, $X_t \sim I(0)$, and unrelated

Figure 30 shows the proportion of times we find a statistically significant long-run effect when the coefficient on x_{t-1} (ARDL/ECM) or x_t (LDV) is also statistically significant. Since the static model does not have a long-run effect, it is omitted. The lagged dependent variable (LDV) model does not include the lag of x_t, but I include it for comparison to the ARDL/ECM specifications (which do). For the LDV model, I proceeded to calculate a long-run effect only if the coefficient on x_t was statistically significant. The results are fascinating, primarily for the fact that we end up with different rejection rates between the ARDL and ECM models. To be clear, the long-run effects are identical, as shown numerous times throughout this SI.[3] However, the strategy of proceeding to calculate the LRM *depending on the significance of the lag of x_t—and then going on to show these proportions*—is what causes the plots to appear different. In Figure 30, we find that the ARDL model produces much lower rates of spurious long-run effects when this strategy is pursued compared to the ECM. The LDV appears relatively unchanged to results found in the main manuscript under this strategy. Note of course that these rejection rates will always be smaller than those shown in Figure 2 in the main manuscript (which showed any significant LRM without regards to the statistical significance of x_{t-1} (ARDL/ECM) or x_t (LDV).

[3]The short-run effects are identical too, as are the rejection rates when examining all simulations (i.e., without the rejection strategy proposed by EMW.

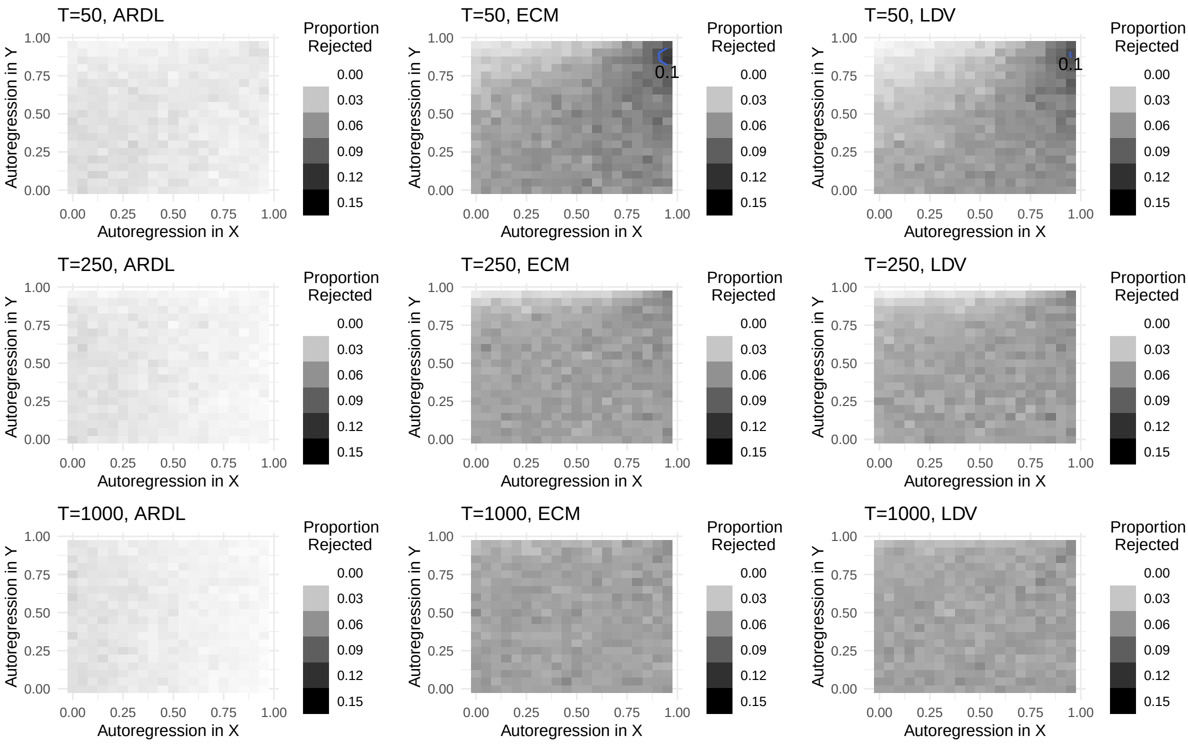

Figure 30: Scenario I: Long-run rejection rates, calculated *only* if x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant

Note: Figure shows the proportion of times the LRM was statistically significant when x_{t-1} (ARDL/ECM) or x_t (LDV) was also statistically significant. Contour lines show boundary of 10 percent rejection rates.

To summarize the findings for Scenario I, when we pursue a strategy of only calculating the LRM if x_{t-1} (ARDL/ECM) or x_t (LDV) is first found to be statistically significant, we are likely to avoid spurious inferences about LRMs *only* for the ARDL model, as compared to always calculating the LRM without regards to the other coefficients.

2.2.2 Scenario II: $Y_t \sim I(0)$, $X_t \sim I(1)$, and unrelated

The results showing the same strategy outlined above for Scenario II ($Y_t \sim I(0)$, $X_t \sim I(1)$, and unrelated) are shown in Figure 31. Similar to the findings in Scenario I, using the strategy of not calculating the LRM if x_{t-1} (ARDL/ECM) or x_t (LDV) is not statistically significant is a good strategy (when series are spuriously related) only for the ARDL specification, with spurious LRMs almost never occurring. In contrast, for both the ECM and LDV specifications, spurious LRMs occur relatively often, and somewhat inconsistently across the level of autoregression in y_t.

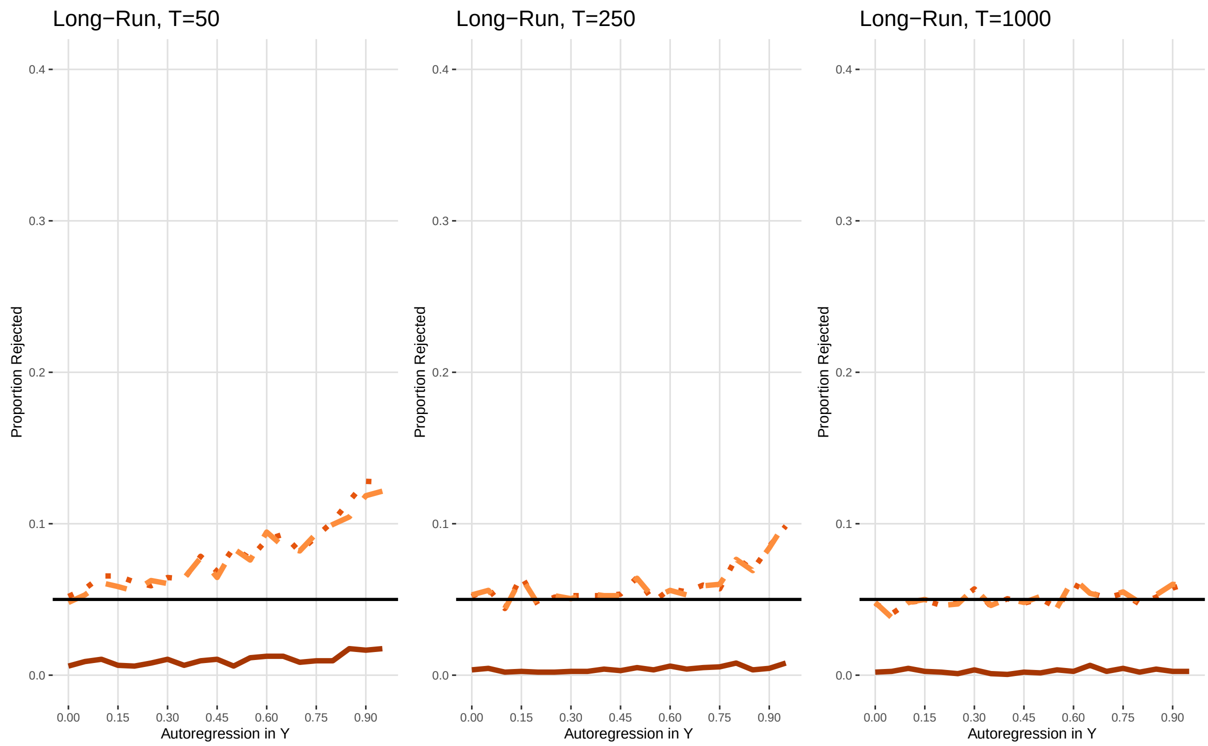

Figure 31: Scenario II: Long-run rejection rates, calculated *only* if x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant

Note: Figure shows the proportion of times the LRM was statistically significant when x_{t-1} (ARDL/ECM) or x_t (LDV) was also statistically significant. Solid line: ARDL, dotted: ECM, dashed: LDV.

Why do the ECM and ARDL models differ so much in Figure 31? In this scenario, the equation balance issue is with the regressor containing a unit root, not the dependent variable. With the ARDL model, by including both x_t and x_{t-1}, the two unit-root regressors are partialing out much of their shared variation (which is extremely high, given that x_t is I(1); really all that is left here is innovations occurring in x between $t-1$ and t). In contrast, in the ECM, only x_{t-1} contains a unit root; Δx_t is made stationary through first-differencing. Thus, there is little shared covariation between x_{t-1} and Δx_t (or y_{t-1}, which is also stationary), which likely gives rise to a large number of times when the coefficient on x_{t-1} is statistically significant. A similar phenomenon occurs with the LDV model, since only y_{t-1} is being used to partial out covariation in x_t. In fact, the two seem to converge at T grows.

2.2.3 Scenario III: $Y_t \sim I(1)$, $X_t \sim I(0)$, and unrelated

I show results of the EMW applied to Scenario III in Figure 32. Consistent with the findings above, the ARDL once again results in very low rates of finding a spurious LRM with this strategy, and spurious LRMs are about equally likely using the ECM or LDV specifications. However, spurious LRMs only occur at very high rates of autoregression in x_t, and decline such that by $T = 250$, no model finds evidence of spurious LRMs above convention.

2.2.4 Scenario IV: $Y_t \sim I(1)$, $X_t \sim I(1)$, and unrelated

For the final spurious scenario, both series contain a unit root. The rejection rates of the LRM using the strategy outlined above are shown in Figure 33. Similar to our earlier findings, the ARDL model has spurious LRMs occur well under convention when using a strategy of only calculating it if x_{t-1} is statistically significant. In contrast, this strategy results in high Type I error for both the ECM and LDV specifications. While rejection rates are smaller than in the main manuscript where this strategy was not used (those rejection rates were around 17%), they are still above convention for the ECM and LDV.

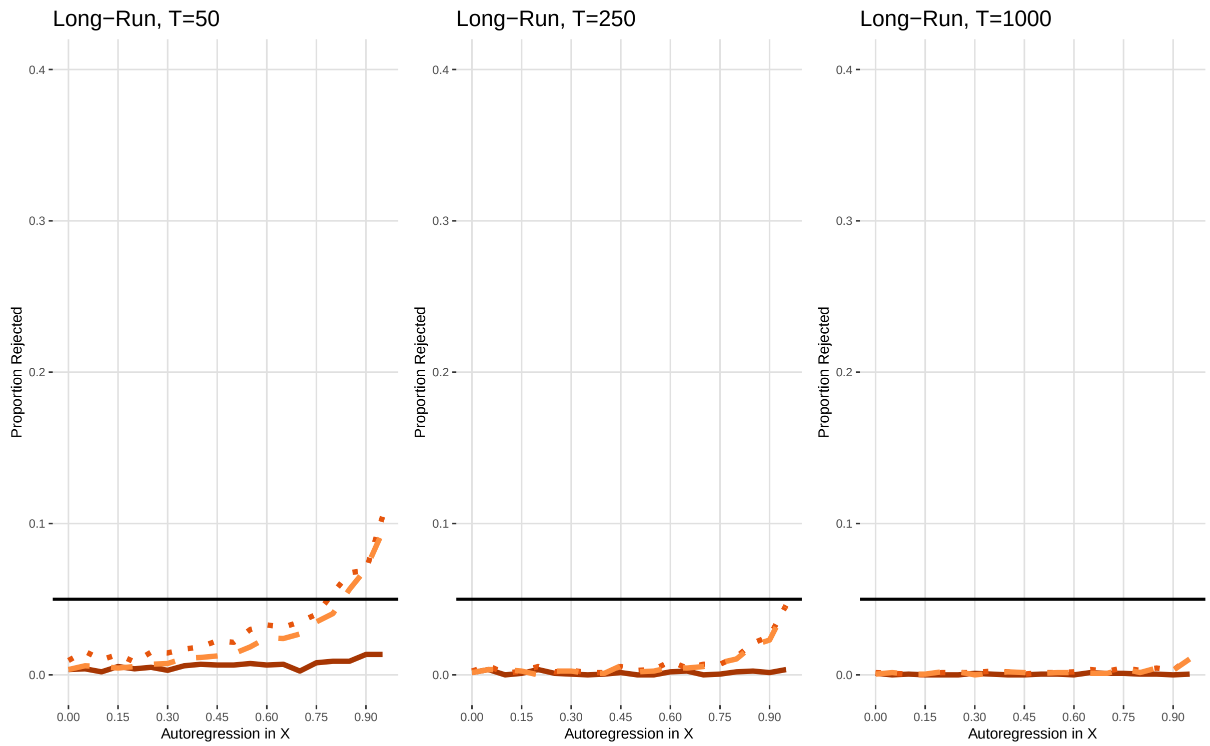

Figure 32: Scenario III: Long-run rejection rates, calculated *only* if x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant

Note: Figure shows the proportion of times the LRM was statistically significant when x_{t-1} (ARDL/ECM) or x_t (LDV) was also statistically significant. Solid line: ARDL, dotted: ECM, dashed: LDV.

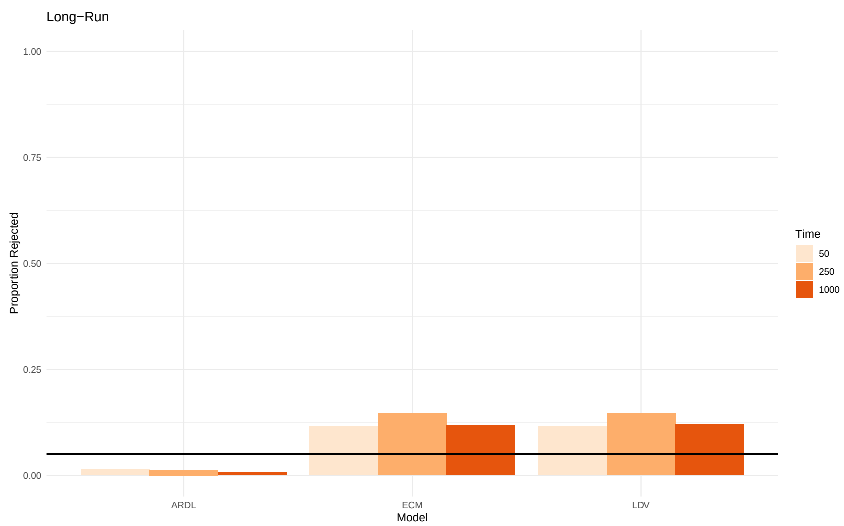

Figure 33: Scenario IV: Long-run rejection rates, calculated *only* if x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant

Note: Figure shows the proportion of times the LRM was statistically significant when x_{t-1} (ARDL/ECM) or x_t (LDV) was also statistically significant.

2.2.5 Scenario V_a: $Y_t \sim I(0)$, $X_t \sim I(0)$, and related

Recall that Scenario V examined true relationships between stationary series. How would the EMW strategy work here? Let:

$$y_t = \alpha y_{t-1} + 2x_t + \beta_2 x_{t-1} + \varepsilon_t \tag{3}$$

where x_t has a contemporaneous effect of $\beta_1 = 2$, $\alpha = 0.2, 0.8$ and $\beta_2 = -1, 0, 1$ and $\varepsilon_t \sim N(0, \sigma_y^2)$ where $\sigma_y^2 = 1, 5$. I include the possibility of $\beta_2 = 0$ in these simulations since EMW's suggestion would lead us to not go on to calculate a long-run effect where $\beta_2 = 0$, even though there would still exist a non-zero LRM of $\frac{\beta_1+\beta_2}{1-\alpha} = \frac{2+0}{1-\alpha}$. Thus, this likely represents a difficult test of their recommendation, even though such a data-generating process—where x_t affects y_t at time t but not one period later—is entirely plausible in social science situations. Unlike Scenario V in the main manuscript, I also allow x_t to be possibly autoregressive of order AR(1), i.e., $x_t = \alpha_x x_{t-1} + u_t$ where $\alpha_x = 0, 0.75$ and $u_t \sim N(0, 1)$.[4] Last, by varying the signal-to-noise ratio, we are able to examine possible changes in standard errors, which again should affect our rejection rates with this strategy.

The results from 2000 simulations using the setup described above are shown in Figure 34. The vertical axis represents the "proportion missed", or the number of simulations for which we failed to find evidence that x_{t-1} (ARDL/ECM) or x_t (LDV) was statistically significant—simulations where the constructed 95% confidence intervals did correctly cover the true LRM—expressed as a proportion of all simulations where the constructed 95% confidence interval correctly encompassed the true long-run effect:[5]

$$\text{Proportion Missed} = \frac{\text{Number of simulations where } \beta_2 = 0 \text{ AND correct LRM found}}{\text{Number of simulations that correctly found LRM}} \tag{4}$$

Essentially, it is the proportion of times that we would have stopped after finding that x_{t-1} (ARDL/ECM) or x_t (LDV) was not statistically significant, and not gone on to

[4] Both u_t and ε_t were generated independently.

[5] Equation 4 uses a test of $\beta_2 = 0$ for the ARDL/ECM, and $\beta_1 = 0$ for the LDV.

calculate the LRM (which would have correctly covered the true LRM); this is exactly the suggestion of EMW. A value of 0 means that no simulations were missed using this approach, while 1 would indicate that 100% of all simulations that would have correctly found evidence of the LRM were missed. All simulations shown in Figure 34 set $\sigma_y^2 = 1$ (i.e., the “low noise” scenario).

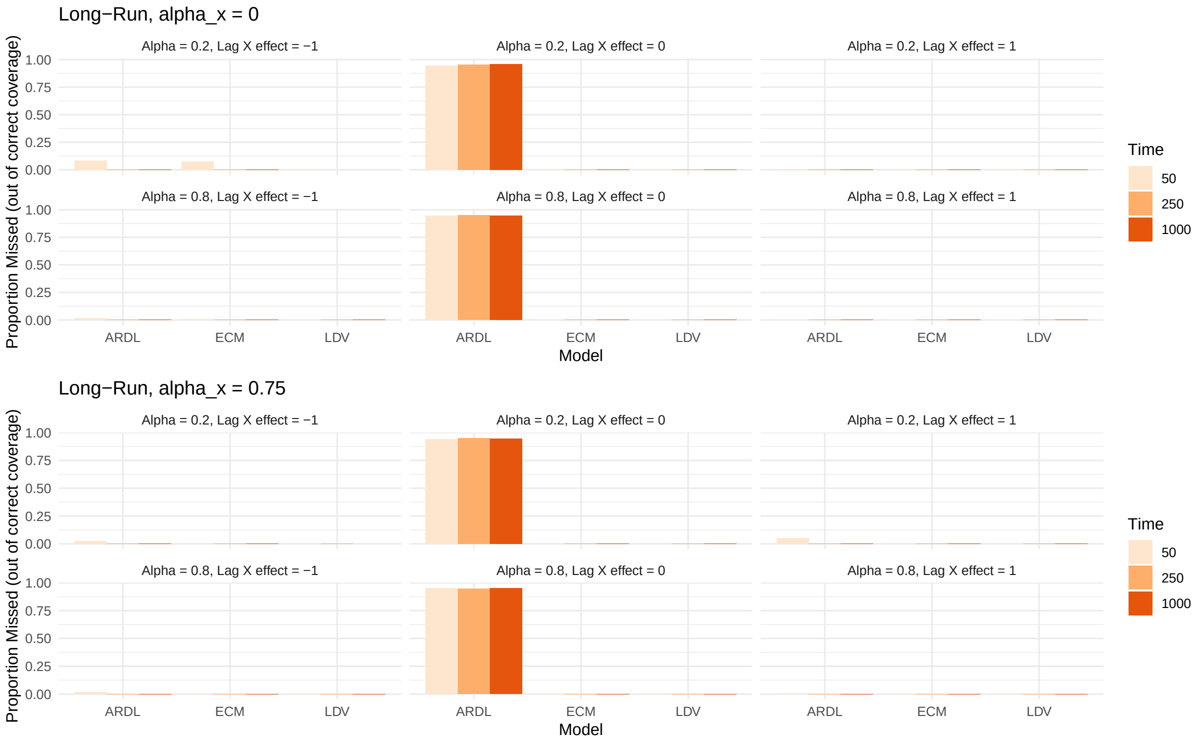

Figure 34: Scenario V_a: Proportion of missed long-run effects if *only* evaluating when x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant, $\sigma_y^2 = 1$

As is clear from Figure 34, we almost always miss out on calculating a (true) long-run effect when following EMW’s suggested strategy when x_t has an effect of zero and when using the ARDL specification. The LDV and ECM specifications do not have this issue. We also miss out in short series; missing around 10 percent of long-run effects we would have correctly calculated when $T = 50$, $\alpha_x = 0$ and we use the ARDL or ECM. Overall, there does not appear to be much of a difference in the results as the level of autoregression in x_t varies.

Figure 35, which shows the same setup as in Figure 34 but raises the error variance in y_t to $\sigma_y^2 = 5$, show much larger issues with the strategy of EMW. Once again the ARDL misses out on calculating what would have been a LRM with correct coverage when the effect of the lag of x is zero. But now issues also arise for both the ARDL and ECM specifications when the lag x_t effect is -1, as well as when it equals 1 (for the ARDL only). These omissions are huge; when x_t is not autoregressive, it's lagged effect is -1, and the level of autoregression is $\alpha_y = 0.2$, we miss out on over *half* of all long-run effects when using the "evaluate x_{t-1} first" strategy, instead of just calculating the long-run effect. Thus, while the EMW strategy appeared to be useful at minimizing Type I error (especially for the ARDL), it drastically reduces our ability to find evidence of long-run effects when they exist.[6]

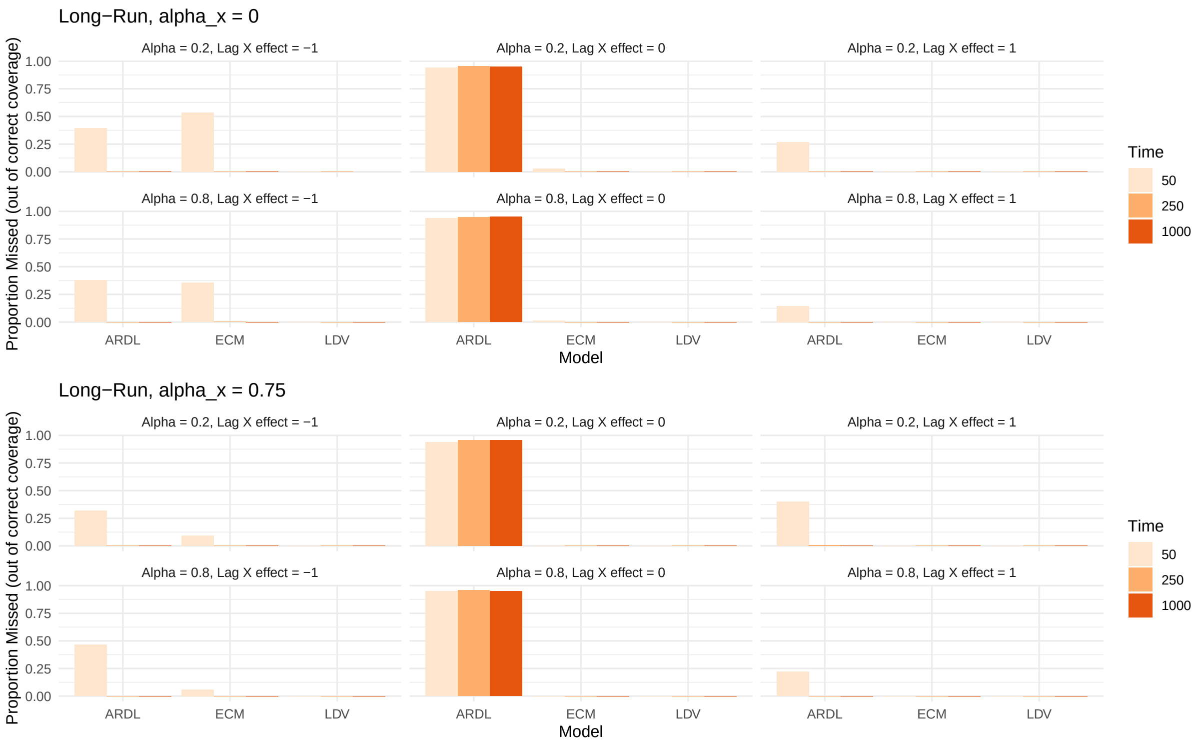

Figure 35: Scenario V_a: Proportion of missed long-run effects if *only* evaluating when x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant, $\sigma_y^2 = 5$

2.2.6 Scenario V_b: $Y_t \sim I(0)$, $X_t \sim I(0)$, $Z_t \sim I(0)$, and related

To add additional, more realistic complexity, consider an additional issues that this strategy may face: additional regressors may be include that might be correlated with x_t. This could make it more likely for multicollinearity to cause us to find x_{t-1} (ARDL/ECM)

[6]Similar, slightly smaller proportions result when we plot these results as a proportion of all simulations.

or x_t (LDV) statistically significant less often.

I set up this scenario as follows. Let y_t be a dependent variable whose data-generating process is:

$$y_t = \alpha y_{t-1} + 2x_t + \beta_2 x_{t-1} + z_t + z_{t-1} + \varepsilon_t \tag{5}$$

where $\alpha = 0.2, 0.8$, $\beta_2 = -1, 0, -1$, $\varepsilon \sim N(0, \sigma_y^2)$ (where $\sigma_y^2 = 1, 5$), and now an additional regressor z_t is introduced with coefficients $\beta_3 = \beta_4 = 1$. x_t and z_t are themselves autoregressive:

$$x_t = 0.5x_{t-1} + u_t \tag{6}$$

$$z_t = 0.5z_{t-1} + \nu_t \tag{7}$$

and also correlated with one another, since $u_t, \nu_t \sim N(0, \Sigma)$ where $\Sigma = \begin{bmatrix} 1 & \sigma_{u,\nu} \\ \sigma_{u,\nu} & 1 \end{bmatrix}$ and $\sigma_{u,\nu} = 0.2, 0.75$. Such a covariance produces a correlation coefficient between x_t and z_t of 0.2 and 0.75 as well.[7] Therefore, we are able to see how having additional correlated regressors may affect this strategy.

The results are presented in a similar manner as Scenario V_b, with $\sigma_y^2 = 1$ in Figure 36 and $\sigma_y^2 = 5$ in Figure 37. The top six plots in each figure are when $\sigma_{u,\nu} = \rho_{x,z} = 0.2$, and the bottom six are when $\sigma_{u,\nu} = \rho_{x,z} = 0.75$. Results when the error variance of y_t (Figure 36) is small are largely the same as those in Figure 34. The largest difference here is when x_t and z_t are highly correlated. Especially when T is small, the EMW strategy will miss out on finding LRMs when using the ARDL, ECM (and to some extent the LDV).

[7]Since $\rho_{x,z} = \frac{\frac{\sigma_{u,\nu}}{1-\alpha_x\alpha_z}}{\sqrt{\frac{\sigma_u^2}{1-\alpha_x^2}}\sqrt{\frac{\sigma_\nu^2}{1-\alpha_z^2}}} = \frac{\frac{\sigma_{u,\nu}}{1-0.5^2}}{\sqrt{\frac{1}{1-0.5^2}}\sqrt{\frac{1}{1-0.5^2}}} = \sigma_{u,\nu}$

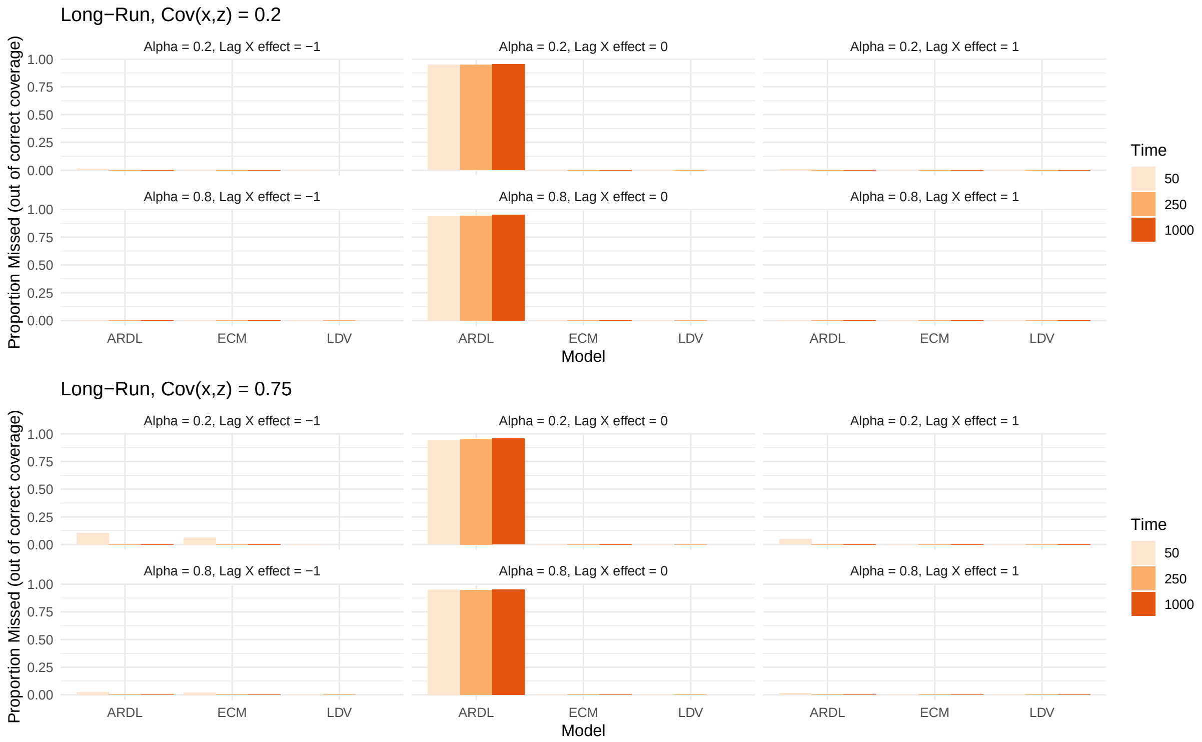

Figure 36: Scenario V_b: Proportion of missed long-run effects if *only* evaluating when x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant with additional regressor, $\sigma_y^2 = 1$

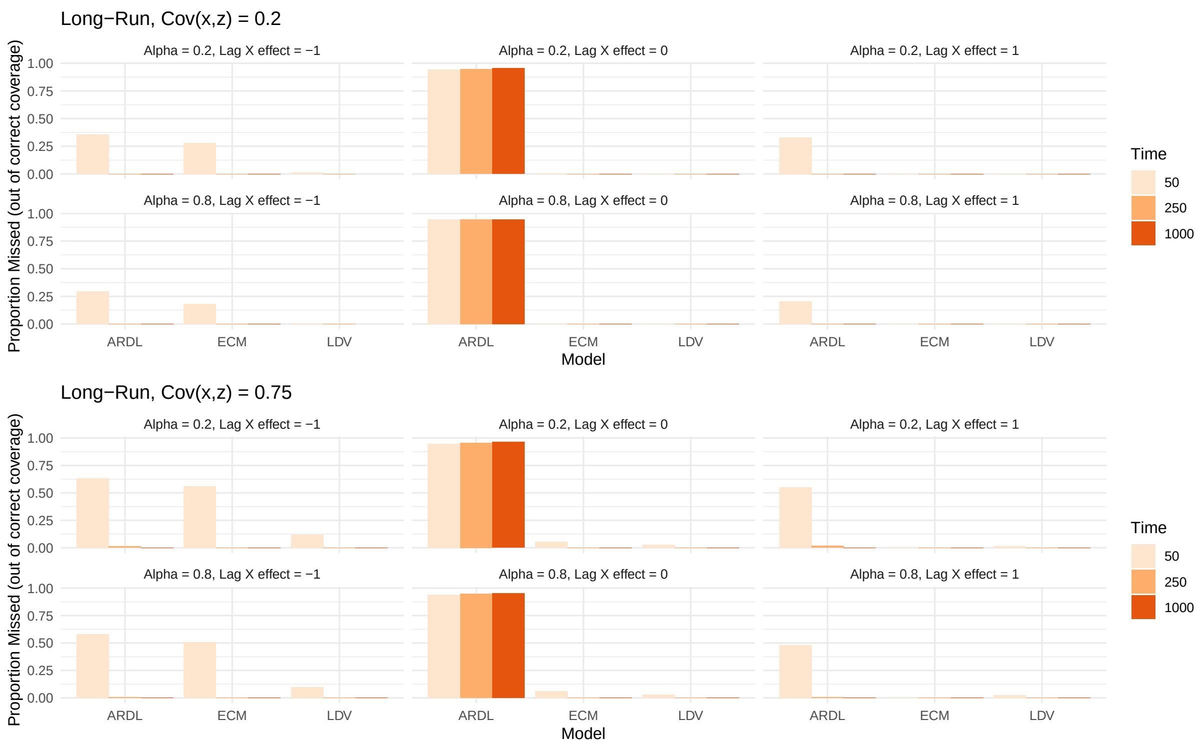

Figure 37: Scenario V_b: Proportion of missed long-run effects if *only* evaluating when x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant with an additional regressor, $\sigma_y^2 = 5$

2.2.7 Scenario VI$_a$: $Y_t \sim I(1)$, $X_t \sim I(1)$, and cointegrating

Turning to examining two cointegrating series, in this sub-section I explore how often we miss out on finding LRMs when we pursue the strategy of not calculating them if we first find that x_{t-1} (ARDL/ECM) or x_t (LDV) are not statistically significant. The data-generating process is the same as in the main manuscript. As above, now I calculate the proportion of missed long-run effects using the EMW strategy.

The results from Scenario VI$_a$ are shown in Figure 38. The top four plots show different combinations of α and the effect of x_{t-1} when $\sigma_y^2 = 1$, and the lower four plots when $\sigma_y^2 = 5$. First, the only time this strategy diverges from just calculating the long-run effect always appears to be when $T = 50$. The LDV appears to suffer from not finding long-run effects when they exist when the effect of $x_{t-1} = -1$. Since it is just estimating a single parameter for x_t—and since the short-run effect of Δx_t is equal to two in the data-generating process, this finding is likely because the positive and negative effects are cancelling each other out. Both the ECM and ARDL models do not miss out on any long-run effects using the proposed strategy except when the effect of $x_{t-1} = 1$, and especially when $\sigma_y^2 = 5$; for instance, when $\alpha = -0.2$, we find only about half of the LRMs that we should have when using the EMW strategy.

2.2.8 Scenario VI$_b$: $Y_t \sim I(1)$, $X_t \sim I(1)$, $Z_t \sim I(1)$, and cointegrating

Similar to the stationary case, it might be that an additional regressor affects the performance of the strategy of not calculating the LRM unless the coefficient on x_{t-1} (ARDL/ECM) or x_t (LDV) are not statistically significant. There appears to be some evidence in the econometric literature of this. For instance, Hendry (1986, p. 207) explains that, "If a model involves I(0) and I(1) variables such that the latter are cointegrated, then there will be a 'near-singularity' in the second moment matrix" meaning that the variables in the cointegrating part of the equation "will be 'highly collinear' and neither need have 'significant "t"-values' despite being cointegrated." If this was the case, we might not find a significant effect until we evaluated the LRM.

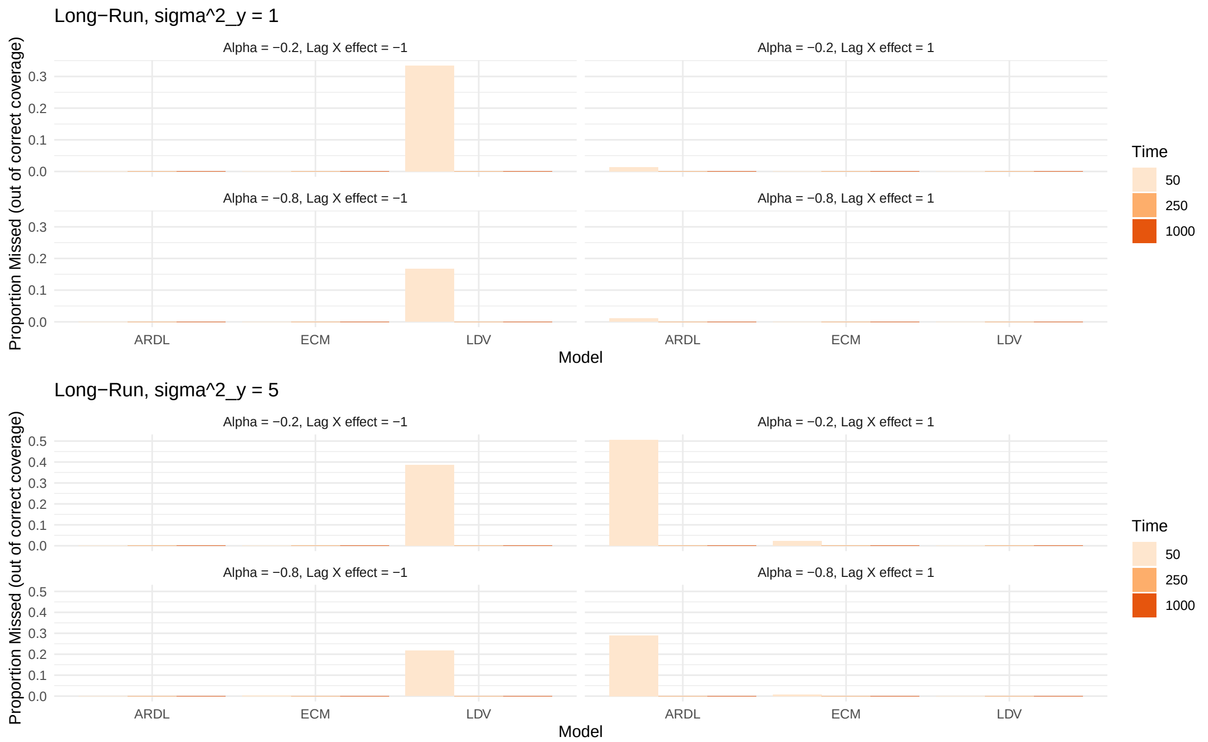

Figure 38: Scenario VI_a: Proportion of missed long-run effects if *only* evaluating when x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant

I set up this scenario the same as the above scenario, except I also introduce a unit-root z_t series. y_t is now generated as:

$$\Delta y_t = \alpha y_t + 2\Delta x_t + \beta_2 x_{t-1} - \Delta z_t + z_{t-1} + \varepsilon_t \tag{8}$$

where $\alpha = -0.2, -0.8$, $\beta_2 = -1, 1$, and both x_t and z_t are unit root series that are unrelated to one another.

The results, using the same setup as in Figure 38, are shown in Figure 39. Perhaps surprisingly, as a result of adding an uncorrelated additional regressor to the equation, overall the proportion of missed long-run effects appears to have decreased for the ARDL model, although risen slightly for the ECM and LDV model when the effect of x_{t-1} is 1. Future work should consider whether cointegrating regressors that are correlated may change these findings.

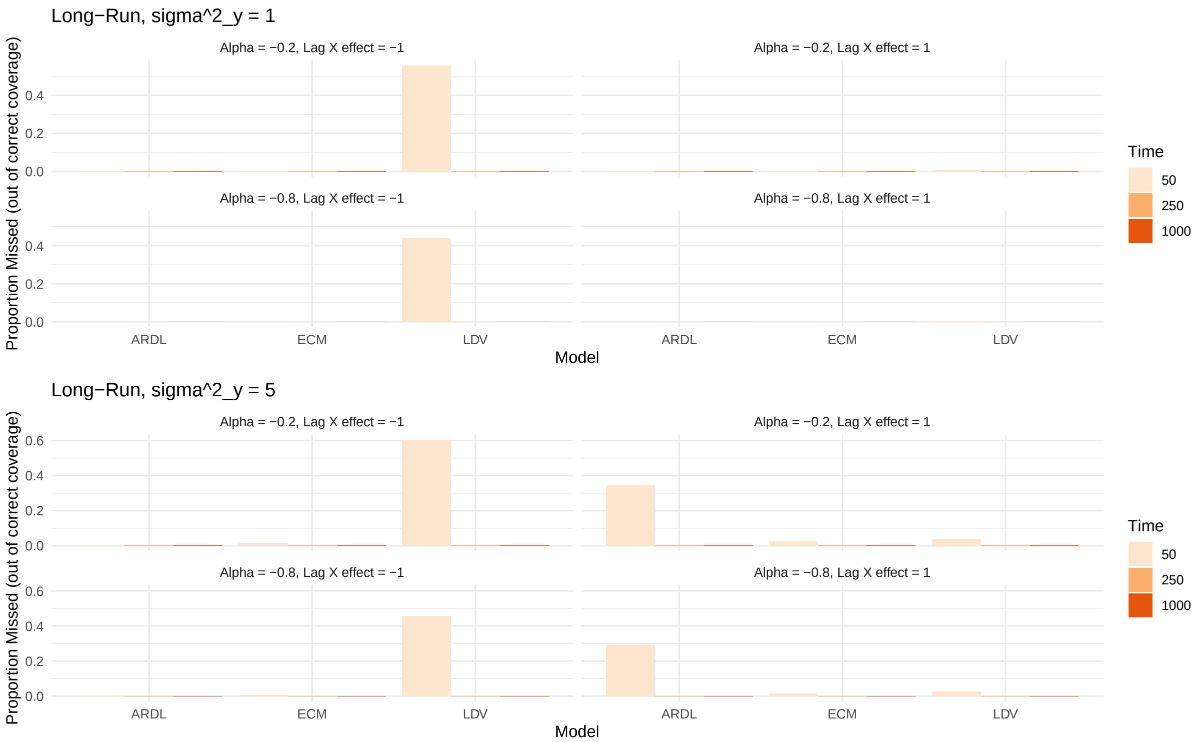

Figure 39: Scenario VI_b: Proportion of missed long-run effects if *only* evaluating when x_{t-1} (ARDL/ECM) or x_t (LDV) is statistically significant with an additional regressor

2.2.9 Conclusion

This section examined whether the strategy of Enns, Moehlecke and Wlezien is a feasible one. This involved first establishing whether x_{t-1} is statistically significant (or x_t for the LDV), and, if so, then going onto calculating long-run effects. A few important findings stand out:

- Results for Scenario I and IV suggest that—especially if the ARDL specification is used—spurious long-run effects are often avoided if one does not calculate the LRM unless x_{t-1} is statistically significant (or x_t for the LDV). Thus, if our goal is to minimize Type I error, this strategy works very well, particularly with the ARDL.

- The issue of course is that minimizing one type of error typically comes at the cost of another. As I have shown, when using the strategy of EMW, we are likely to miss out on calculating long-run effects when they *do* exist. The very specification that performed the best in terms of Type I error—the ARDL—had the largest issues in regards to this.

- The EMW strategy often leads to different conclusions based on whether an ARDL or ECM specification is estimated. This is interesting, given that in terms of short- and long-run effects, the two are identical (De Boef and Keele 2008). It is also makes this strategy awkward. Should we use the ARDL model or ECM, given that although they provide the exact same short- and long-run effects, we might end up not calculating long-run effects more often with the former than the latter?

- That we should always be evaluating x_{t-1} before going onto estimate a long-run effect is rather silly; as shown in Scenario V$_a$ and Scenario V$_b$, in stationary relationships it is entirely possible for the contemporaneous effect of x_t to be statistically significant while x_{t-1} has no effect, and the EMW strategy would almost always fail to go onto calculate long-run effects when they exist.

- The findings that were most in favor of the EMW strategy were the results for cointegrating series (Scenario VI$_a$ and VI$_b$). Even so, the question remains; why not just test for cointegration in the first place? If the analyst finds evidence for this, they can proceed to calculate long-run effects since, by definition, they should exist. One possible drawback here is that cointegration tests tend to be "all in" when there are multiple regressors; either there is evidence of cointegration—for which we conclude that *all* I(1) regressors are cointegrating (and thus can proceed to calculating long-run effects), or *none* are (thus, we do not calculate long-run effects and need to respecify the model). Yet even if we conclude cointegration, it might still be the case that one individual regressor is not cointegrating and the test just lacks the power to detect it. While approaches to parsing out individual tests exist—such as an inconclusive 'bounds' critical value (Pesaran, Shin and Smith 2001; Philips 2018)—or bounds on individual effects themselves (Webb, Linn and Lebo 2019, 2020), they are still relatively underutilized or are new approaches.

- In addition, as we know from cointegration, ECMs—as political science commonly estimates them—"hide" the long-run effect. It is a combination of the lag of the regressors and the adjustment parameter. While the latter is nearly always statistically significant, as shown in the Monte Carlos above, the former might not be. Indeed, this is some of the reason as to why many cointegration tests involve the *joint* significance of the regressors and the lag of y_t (Pesaran, Shin and Smith 2001; Philips 2018) . While we should not conflate cointegration with an individual regressor necessarily having a long-run effect, cointegration implies that at least some of the regressors do.

- The signal-to-noise ratio appears to be a strong driver of whether the EMW strategy results in failing to calculate a LRM when one should. Higher noise—as proxied in the Monte Carlos by increased variance in the residual of y_t—appears to always make the EMW strategy worse.

- The EMW strategy works quite well when $T = 250$ or greater. Issues, which tended to be very large in Scenarios V and VI, only really arose when $T = 50$. Of course, time series in political science applications tend to be relatively short.

The proposal of EMW is interesting, and certainly worthy of further research. Still, a conditional strategy of testing for the significance of x_{t-1} (or x_t for the LDV) in order to determine whether users should move onto calculate a long-run effect seems to be overall unwise, as I have shown using the Monte Carlos above. Instead, scholars should first establish stationarity (and possibly cointegration) conditions, then move on to testing for short- and long-run effects.

References

De Boef, Suzanna and Jim Granato. 1997. "Near-integrated data and the analysis of political relationships." *American Journal of Political Science* pp. 619–640.

De Boef, Suzanna and Luke Keele. 2008. "Taking time seriously." *American Journal of Political Science* 52(1):184–200.

Enders, Walter. 2010. *Applied econometric time series.* 3 ed. John Wiley and Sons.

Engle, Robert F and Clive WJ Granger. 1987. “Co-integration and error correction: representation, estimation, and testing.” *Econometrica* pp. 251–276.

Granger, Clive WJ and Paul Newbold. 1974. “Spurious regressions in econometrics.” *Journal of Econometrics* 2(2):111–120.

Hendry, David F. 1986. “Econometric modelling with cointegrated variables: an overview.” *Oxford bulletin of economics and statistics* 48(3):201–212.

Komashko, Oleh. 2016. “nlWaldTest: Wald test of nonlinear restrictions and nonlinear CI.” *R package version* 1.1.3.

Pesaran, M Hashem, Yongcheol Shin and Richard J Smith. 2001. “Bounds testing approaches to the analysis of level relationships.” *Journal of Applied Econometrics* 16(3):289–326.

Philips, Andrew Q. 2018. “Have your cake and eat it too? Cointegration and dynamic inference from autoregressive distributed lag models.” *American Journal of Political Science* 62(1):230–244.

Pickup, Mark. 2014. *Introduction to time series analysis*. Vol. 174 Sage Publications.

Webb, Clayton, Suzanna Linn and Matthew J Lebo. 2020. “Beyond the Unit Root Question: Uncertainty and Inference.” *American Journal of Political Science* 64(2):275–292.

Webb, Clayton, Suzanna Linn and Matthew Lebo. 2019. “A Bounds Approach to Inference Using the Long Run Multiplier.” *Political Analysis* 27(3):281–301.