But Wait, There's More! Maximizing Substantive Inferences from TSCS Models

Online Appendix

Laron K. Williams

Department of Political Science

University of Missouri

and

Guy D. Whitten

Department of Political Science

Texas A&M University

Overview

This document provides analyses briefly discussed—but not presented—in "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models".

Forecast Errors

In this section, we present evidence that generating forecast errors through simulation methods will provide the same substantive inferences as those generated via analytical methods. Regardless of the $N \times T$ configuration, the degree of the autoregression, and the model fit, dynamic simulations with forecast errors will echo those inferences made with analytically-derived forecast errors.

Enders (2004) provides the formula for the conditional variance of the forecast. First, assume that we estimate the following model: $y_t = \alpha_0 + \alpha_1 y_{t-1} + \epsilon_t$. The conditional forecast at time $t+1$ is $E_t y_{t+1} = \alpha_0 + \alpha_1 y_t$. If we assume that the elements of the e_t sequence are independent and have a variance of σ^2, then the variance of the j-step-ahead forecast error is $Var[e_t(j)] = \sigma^2[1 + \alpha_1^2 + \alpha_1^4 + \alpha_1^6 + \cdots + \alpha_1^{2(j-1)}]$. Since the size of the forecast error increases as a function of j, we will have greater confidence in short-term forecasts than long-term forecasts. Also important to note is that "in the limit as $j \to \infty$, the forecast error variance converges to $\sigma^2/(1 - \alpha_1^2)$" (Enders 2004: 80).

We can derive the conditional variance of the forecast error analytically by calculating the σ^2 and the α_1 from our model. Or, we can make use of the asymptotic sampling distribution from Clarify to generate a large number of replicates ($r = 1000$) of both σ^2 and α_1, and then use the percentile method to calculate confidence intervals. We prefer to use simulation-based methods over analytical methods, a preference that is echoed by King, Tomz and Wittenberg (2000: 352-353).

To demonstrate the similarities of these two approaches, we present simulation evidence and figures from the following model estimated on varying sample sizes (N and T): $y_t = 0.25 + (\alpha_1 \times y_{t-1})$, where we modify the values of α_1. Figures S.1 and S.2 demonstrate the similarity between analytically-derived and simulation-based confidence intervals across degree of trend (where each of the four cells represents increasing values of α), as well as sample size. We slightly jitter the confidence intervals so that they are distinguishable.

[Figures S.1 and S.2 about here]

At the same configurations of N and T, as well as level of autoregression (α_1) the analytically-derived and simulation-based forecasting errors give virtually identically-sized confidence intervals. While we only show simulation evidence for two sample configurations (i.e., 10×10 and 20×40), it is our experience that scholars can be confident in using simulation-based methods to produce dynamic simulations with forecast errors that are virtually identical to those derived analytically regardless of the sample size or model fit.

Simulation-Based Coverage Rates

In this section we explore the credibility of the multivariate normal assumption underpinning simulation methods to sample sizes traditionally used in TSCS. These Monte Carlo experiments represent a worthwhile endeavor since Beck and Katz's (1995) focus on comparing panel-corrected standard errors to the Parks method leaves us to wonder how OLS performs under these conditions. We follow Beck and Katz's (1995) lead in calculating the *level* and *overconfidence* as a way of assessing OLS regression's performance. While *level* represents the proportion of the replicates where the confidence interval contains the true coefficient, we calculate *overconfidence* the following way:

$$Overconfidence = 100\frac{\sqrt{\sum_{l=1}^{1,000}(\tilde{\beta}^{(l)} - \bar{\tilde{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000}(s.e.(\tilde{\beta}^{(l)}))^2}},$$

where $\tilde{\beta}^{(l)}$ is the OLS estimate of β_1 for replication l. *Overconfidence* measures the extent to which the estimator understates variability. If the estimator produces standard errors that accurately reflect its variability, then *overconfidence* is 100. And, if the estimator produces standard errors that overstate the true sampling variability, and thus the standard errors are too large, *overconfidence* is less than 100 (Beck and Katz 1995: 639).

In each experiment, we varied N $(i = 1, \ldots, N)$ and T $(t = 1, \ldots, T)$ to mimic traditional sample size configurations found in a survey of TSCS models (Beck and Katz 1995: 635). We then generated 1000 replicates of the NT error terms according to the model being specified, $\epsilon_{i,t}^{(l)}$, where $i = 1, \ldots, N$, $t = 1, \ldots, T$ and $l = 1, \ldots, 1000$. For each replicate, l, the errors are generated as a zero-mean NT-variate normal distribution with the covariance structure varying depending on the panel error assumptions. We fixed the $x_{i,t}$ over the 1000 replicates and generated 1000 replicates of $y_{i,t}$ using

$$y_{i,t}^{(l)} = 0.25 + 10 \times x_{i,t} + \epsilon_{i,t}^{(l)}; i = 1, \ldots, N; t = 1, \ldots, T; l = 1, \ldots, 1000.$$

Since we are principally concerned with the performance of OLS standard errors, Table S.1 shows the performance of OLS analytically-derived standard errors (labeled OLS-A) to those calculated via simulation methods (labeled OLS-S). We calculate 1000 replicates of a sample with an $N = 10$, varying levels of T, as well as contemporaneous correlation. We modify the variance covariance matrix of the errors so that all non-contemporaneous observations are 0 and the values of the contemporaneous correlation varies. For example, if $N = 2$ and $T = 3$, the variance covariance matrix of the errors is the following (Beck and Katz 1995: 646):

$$\mathbf{\Omega} = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \sigma_{12} & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 & \sigma_{12} & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & \sigma_{12} \\ \sigma_{12} & 0 & 0 & \sigma_2^2 & 0 & 0 \\ 0 & \sigma_{12} & 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{12} & 0 & 0 & \sigma_2^2 \end{pmatrix}$$

We will first explore instances of contemporaneous correlation. Tables S.1-S.3 show how OLS analytically-derived standard errors compare to simulation-based standard errors in terms of *overconfidence* and *level*. Table S.1 has an $N = 10$, whereas Table S.2 and Table S.3 have N of 15 and 20, respectively.

[Tables S.1-S.3 about here]

Under all conditions of contemporaneous correlation and sample configuration, OLS standard errors perform quite well. In all circumstances, the *overconfidence* is close to 100, and never strays either below 95 or above 105. Likewise, the *level* is always close to 95 indicating that the 95% confidence interval contains the true β around 95% of the time. This is the case in samples that would be considered small in TSCS analyses ($N = 10$, $T = 10$) as well as large samples ($N = 20$, $T = 40$). Most importantly, the simulation-based standard errors perform nearly identical to those derive analytically under all the different scenarios.

To gauge OLS' performance in times of panel heteroskedasticity, we design experiments with panel-heteroskedastic error structures that are related to the panel structure of the independent variables. For each value of t, we generate the N-vector $x_{i,t}(i = 1, \ldots, N)$ as a draw from a zero-mean N-variate normal distribution. We then set the variance of the first half of the units to 1 while we experimentally manipulate the variance of the second half of the units. Beck and Katz (1995) introduce the following measure of panel heteroskedasticity, which we also use: "the variance of the ith unit is σ_i^2. Let $\omega_i = \frac{1}{\sigma_i}$. We define *standardized heteroskedasticity* as the standard deviation of the $\frac{\omega_I}{\bar{\omega}}$" (646). We can then calculate varying levels of both contemporaneous correlation and panel heteroskedasticity.

In Table S.4 we provide the results of Monte Carlo experiments of varying T (N is fixed at 15 since we found no substantive differences for other values of N), panel heteroskedasticity and contemporaneous correlation.

[Table S.4 about here]

As shown in the previous tables, in those situations with only contemporaneous correlation (i.e., panel heteroskedasticity = 0), both OLS standard errors perform quite well. On the other hand, in the presence of moderate levels of panel heteroskedasticity (0.3), both analytically-derived and simulation-based standard errors tend to dramatically overestimate the size of the standard errors (producing low values of *overconfidence*), which produces confidence intervals that always include the true β (*level*=100). Also important to note is that, even though these standard errors are quite large, both versions of the standard errors (OLS-A and OLS-S) perform similarly.

The Monte Carlo experiments in Table S.4 suggest that using OLS in the presence of panel heteroskedasticity should be strongly discouraged. While the results certainly support this conclusion, we should emphasize that these conditions are unlikely to be replicated in typical research situations. Keep in mind that standard errors are accurate in the presence of panel heteroskedasticity "if the terms in the error covariance matrix, $\mathbf{\Omega}$, are not related to the squares and cross products of the independent variables" (Beck and Katz 1995: 640). As an attempt to demonstrate the conditions under which OLS may not be appropriate, we deliberately manipulate the experiments to extreme conditions—in a sense, "stacking the deck". Indeed, in our experiments the variances and covariances of the errors were proportional to the variances and covariances of the independent variables. It is our belief that there are few real-world situations where the data would be generated in such a way, and to such a high degree of panel heteroskedasticity.

As an effort to expand the capability of our `dynsim` command, and help scholars produce reliable estimates of the variance-covariance matrices, we modify the Stata command to allow estimates from panel-corrected standard errors. We feel that this is extremely helpful, as panel-corrected standard errors perform much better than OLS in terms of level and overconfidence in those situations where OLS performs poorly.

[Table S.5 about here]

Table S.5 compares simulation-based OLS standard errors to those simulation-based panel-corrected standard errors (labeled PCSE-S) for varying levels of T (N is fixed at 15) and contemporaneous correlation in the presence of panel heteroskedasticity (0.3). While still producing standard errors that are larger than those from the true sampling variability (as well as much larger confidence intervals), panel-corrected standard errors offer a drastic improvement over simulation-based OLS standard errors.

Finally, we relax the assumption that the serial correlation follows the same pattern for all units. We calculate the errors according to a unit-specific first-order autoregressive (AR1) process:

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + \nu_{i,t}.$$

The ρ_1 for the first half of the units is 0.9 while we vary the ρ_2 for the second half of the units. To assess the performance of OLS in situations of explanatory variables with varying trends, we also experimentally manipulate the the independent variable:

$$x_{i,t} = \delta x_{i,t-1} + \mu_{i,t}.$$

Table S.6 compares analytically-derived OLS standard errors to those simulation-based OLS standard errors for varying levels of T (N is fixed at 15), panel autocorrelation (ρ_2), and trend (δ). The results should reassure scholars who use OLS in the presence of an autoregressive series or trending independent variables. Under all conditions typically observed in TSCS data, OLS standard errors—either analytically-derived or simulation-based—will have standard errors that largely reflect the true sampling variability and will recover the true β about 95% of the time. Moreover, the similarity of analytically-derived and simulation-based standard errors points to the usefulness of simulation-based methods.

Nevertheless, the caveats of Clarify also apply to our `dynsim` command. As King, Tomz and Wittenberg (2000: 351) state, "we assume that the statistical model is identified and correctly specified[...]which allows us to focus on interpreting and presenting the final results[...]In short, our algorithms work whenever the usual assumptions work." Thus, the impetus is on the scholar. However, in most situations OLS performs surprisingly well, and certainly the simulation-based techniques perform as well as the analytically-derived standard errors. In those rare circumstances where OLS performs poorly, one can use other methods such as panel-corrected standard errors. Given these sample structures and the presence of these problems, we are confident that the general principle of graphing dynamic simulations is broadly applicable.

dynsim: A Program for Producing Dynamic Simulations

In this section we provide some background for using simulation methods to interpret long-term dynamic relationships, introduce our Stata command for producing dynamic simulations, and demonstrate how we use our command to create the figures in the manuscript.

The `dynsim` command is somewhat limited in the number of models that it can run by two constraints: first, the requirement that the dependent variable is continuous, and second, the few models applicable with Clarify. This means that the command only currently produces dynamic simulations for OLS models and, after some slight modifications to the original `estsimp` and `simqi` commands, OLS with panel-corrected standard errors. It is important to note, however, that the idea of sampling from the asymptotic sampling distribution as a means of producing dynamic simulations is one that applies to a broader class of models with continuous dependent variables (for a Bayesian perspective, see Jackman 2009).

For more information, please see "`dynsim`: A Stata Command for Creating Dynamic Simulations of Autoregressive Relationships," *the Stata Journal* forthcoming.

Tables and Figures

Table S.1: Performance of OLS Analytically-Derived and Simulation-Based Standard Errors in the Presence of Contemporaneous Correlation: $N = 10$

		OLS-A		OLS-S	
	CC	Overconfidence (%)[b]	Level[c]	Overconfidence (%)	Level
$T^a = 10$	0.25	101.0	94.8	100.9	94.9
	0.50	101.1	94.6	101.0	94.4
	0.75	101.2	94.6	101.2	97.4
$T = 20$	0.25	100.2	94.2	100.2	94.2
	0.50	100.2	94.1	100.2	93.9
	0.75	100.2	94.3	100.0	93.9
$T = 30$	0.25	100.7	94.1	100.8	93.8
	0.50	100.7	93.9	100.8	93.8
	0.75	100.7	93.8	100.8	93.6
$T = 40$	0.25	98.0	94.6	98.0	94.5
	0.50	98.0	94.6	98.1	94.6
	0.75	98.0	94.7	97.9	94.7

[a] Number of time points.

[b] Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1,000}(\hat{\beta}^{(l)} - \bar{\hat{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000}(s.e.(\hat{\beta}^{(l)}))^2}}$.

[c] Percentage of 95% confidence intervals containing β.

Table S.2: Performance of OLS Analytically-Derived and Simulation-Based Standard Errors in the Presence of Contemporaneous Correlation: $N = 15$

		OLS-A		OLS-S	
	CC	Overconfidence (%)[b]	Level[c]	Overconfidence (%)	Level
	0.25	97.6	96.0	97.6	95.8
$T^a = 15$	0.50	97.7	95.9	97.7	96.0
	0.75	97.6	96.0	97.5	96.0
	0.25	95.0	95.7	95.0	95.7
$T = 20$	0.50	95.1	95.7	95.1	95.3
	0.75	95.1	95.6	95.0	95.5
	0.25	96.2	96.1	96.3	95.8
$T = 30$	0.50	96.3	96.1	96.3	95.9
	0.75	96.3	96.1	96.4	96.3
	0.25	98.9	95.3	98.9	95.0
$T = 40$	0.50	98.8	95.3	98.9	95.0
	0.75	98.9	95.2	98.9	95.0

[a] Number of time points.

[b] Overconfidence $= 100 \frac{\sqrt{\sum_{l=1}^{1,000} (\hat{\beta}^{(l)} - \bar{\hat{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000} (s.e.(\hat{\beta}^{(l)}))^2}}$.

[c] Percentage of 95% confidence intervals containing β.

Table S.3: Performance of OLS Analytically-Derived and Simulation-Based Standard Errors in the Presence of Contemporaneous Correlation: $N = 20$

		OLS-A		OLS-S	
	CC	Overconfidence (%)[b]	Level[c]	Overconfidence (%)	Level
$T^a = 20$	0.25	98.7	95.2	98.8	95.0
	0.50	98.8	94.9	98.8	94.9
	0.75	98.8	94.9	98.9	95.0
$T = 25$	0.25	103.8	95.3	103.8	95.0
	0.50	103.8	95.3	104.0	94.9
	0.75	104.0	95.1	104.0	95.0
$T = 30$	0.25	97.0	95.1	96.9	95.0
	0.50	97.0	95.0	96.9	95.0
	0.75	96.9	95.0	96.9	95.2
$T = 40$	0.25	97.4	95.9	97.4	95.5
	0.50	97.5	95.8	97.4	96.1
	0.75	97.5	96.0	97.5	96.0

[a] Number of time points.

[b] Overconfidence $= 100 \frac{\sqrt{\sum_{l=1}^{1,000}(\hat{\beta}^{(l)} - \bar{\hat{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000}(s.e.(\hat{\beta}^{(l)}))^2}}$.

[c] Percentage of 95% confidence intervals containing β.

Table S.4: Performance of OLS Analytically-Derived and Simulation-Based Standard Errors in the Presence of Contemporaneous Correlation and Panel Heteroskedasticity

			OLS-A		OLS-S	
	PH	CC	Overconfidence (%)[b]	Level[c]	Overconfidence (%)	Level
	0	0	100.4	96.1	100.6	95.5
	0	0.25	100.1	95.8	100.0	95.6
$T^a = 10$	0.3	0	22.4	100.0	22.3	100.0
	0.3	0.25	22.0	100.0	22.0	100.0
	0.3	0.50	22.0	100.0	22.0	100.0
	0	0	94.4	95.5	94.4	95.5
	0	0.25	95.0	95.7	95.0	95.4
$T = 20$	0.3	0	18.3	100.0	18.3	100.0
	0.3	0.25	17.8	100.0	17.8	100.0
	0.3	0.50	17.8	100.0	17.8	100.0
	0	0	97.0	95.6	97.0	95.5
	0	0.50	96.3	96.1	96.2	95.8
$T = 30$	0.3	0	16.0	100.0	16.0	100.0
	0.3	0.25	15.3	100.0	15.3	100.0
	0.3	0.50	15.4	100.0	15.4	100.0
	0	0	99.5	95.6	99.4	95.8
	0	0.50	98.8	95.3	98.8	95.1
$T = 40$	0.3	0	14.2	100.0	14.2	100.0
	0.3	0.25	13.3	100.0	13.3	100.0
	0.3	0.50	13.3	100.0	13.3	100.0

[a] Number of time points; number of units fixed at 15.

[b] Overconfidence $= 100\frac{\sqrt{\sum_{l=1}^{1,000}(\hat{\beta}^{(l)} - \bar{\hat{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000}(s.e.(\hat{\beta}^{(l)}))^2}}$.

[c] Percentage of 95% confidence intervals containing β.

Table S.5: Performance of OLS Analytically-Derived and Simulation-Based Standard Errors in the Presence of Contemporaneous Correlation and Panel Heteroskedasticity versus Panel-Corrected Standard Errors

			OLS-S		PCSE-S	
	PH	CC	Overconfidence (%)[b]	Level[c]	Overconfidence (%)	Level
	0.3	0	22.3	100.0	76.5	99.8
$T^a = 10$	0.3	0.25	22.0	100.0	75.2	99.9
	0.3	0.50	22.0	100.0	75.2	99.9
	0.3	0	18.3	100.0	80.0	99.4
$T = 20$	0.3	0.25	17.8	100.0	78.7	99.7
	0.3	0.50	17.8	100.0	78.4	99.6
	0.3	0	16.0	100.0	82.3	99.2
$T = 30$	0.3	0.25	15.3	100.0	80.5	99.6
	0.3	0.50	15.4	100.0	80.5	99.6
	0.3	0	14.2	100.0	81.0	99.1
$T = 40$	0.3	0.25	13.3	100.0	78.0	99.2
	0.3	0.50	13.3	100.0	78.0	99.2

[a] Number of time points; number of units fixed at 15.

[b] Overconfidence $= 100 \frac{\sqrt{\sum_{l=1}^{1,000}(\hat{\beta}^{(l)} - \bar{\hat{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000}(s.e.(\hat{\beta}^{(l)}))^2}}$.

[c] Percentage of 95% confidence intervals containing β.

Table S.6: Performance of OLS Analytically-Derived and Simulation-Based Standard Errors in the Presence of Panel-Specific Autocorrelation

			OLS-A		OLS-S	
	ρ_2^a	δ^b	Overconfidence (%)[c]	Level[d]	Overconfidence (%)	Level
$T^e = 10$	0.30	0.30	102.0	94.3	101.8	94.5
	0.50	0.30	102.2	94.1	102.3	94.2
	0.30	0.50	102.0	94.3	102.0	93.8
	0.50	0.50	102.2	94.1	102.3	93.6
	0.30	0.90	102.0	94.3	101.9	94.3
	0.50	0.90	102.2	94.1	102.2	93.3
$T = 20$	0.30	0.30	97.9	94.3	98.0	94.3
	0.50	0.30	99.5	94.8	99.5	94.6
	0.30	0.50	98.0	94.3	97.9	94.4
	0.50	0.50	99.5	94.8	99.5	94.1
	0.30	0.90	97.9	94.3	98.0	94.3
	0.50	0.90	99.5	94.8	99.4	94.7
$T = 30$	0.30	0.30	98.2	94.4	98.1	94.1
	0.50	0.30	99.8	94.8	99.7	94.7
	0.30	0.50	98.2	94.4	98.1	94.2
	0.50	0.50	99.8	94.8	99.8	94.3
	0.30	0.90	98.2	94.4	98.2	94.5
	0.50	0.90	99.8	94.8	99.8	94.4
$T = 40$	0.30	0.30	101.3	94.5	101.3	95.0
	0.50	0.30	102.4	93.8	102.3	93.8
	0.30	0.50	101.3	94.5	101.3	94.2
	0.50	0.50	102.4	93.8	102.4	93.8
	0.30	0.90	101.3	94.5	101.4	94.4
	0.50	0.90	102.4	93.8	102.4	93.9

[a] Serial correlation of errors of second half of units; serial correlation of first half is 0.90.

[b] $x_{i,t} = \delta x_{i,t-1} + \mu_{i,t}$.

[c] Overconfidence $= 100 \frac{\sqrt{\sum_{l=1}^{1,000}(\hat{\beta}^{(l)} - \bar{\hat{\beta}})^2}}{\sqrt{\sum_{l=1}^{1,000}(s.e.(\hat{\beta}^{(l)}))^2}}$.

[d] Percentage of 95% confidence intervals containing β.

[e] Number of time points; number of units fixed at 15.

Figure S.1: Analytically-Derived and Simulation-Based Forecasting Errors across Varying Levels of α: $N = 10$, $T = 10$, $\sigma^2 = 0.93$

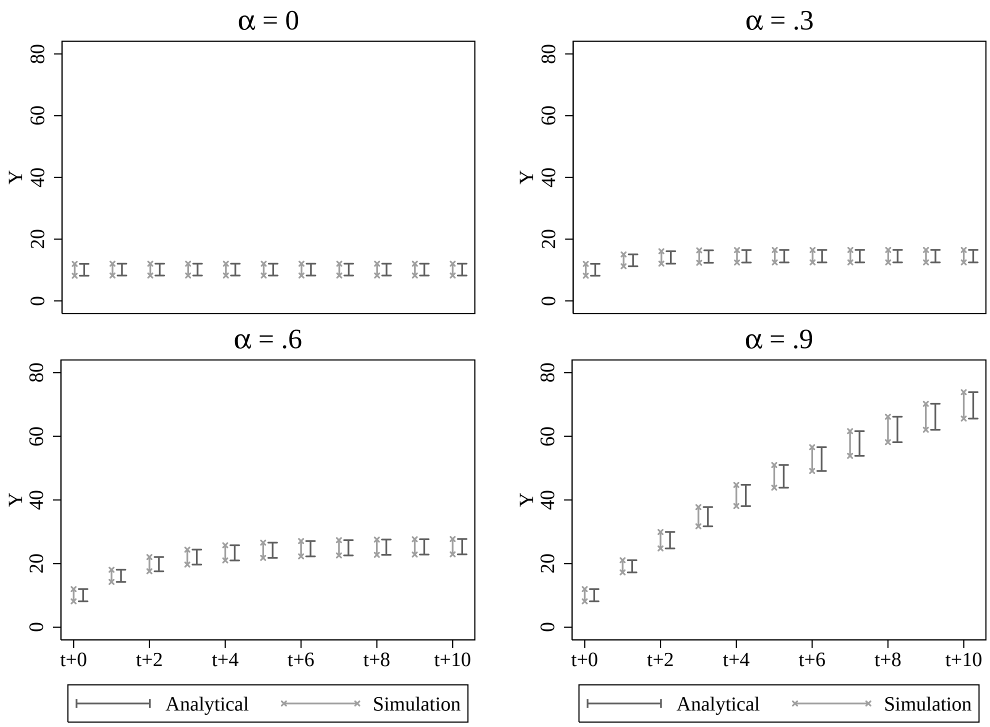
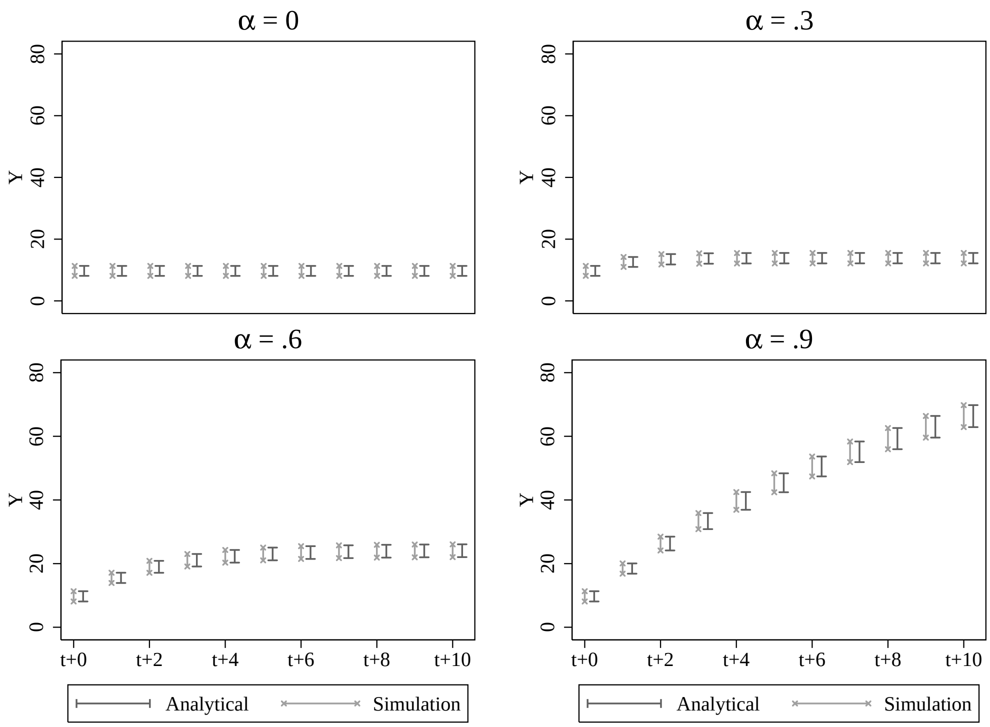


Figure S.2: Analytically-Derived and Simulation-Based Forecasting Errors across Varying Levels of α: $N = 20$, $T = 40$, $\sigma^2 = 0.68$



References

deBoef, Suzanna & Luke Keele. 2008. “Taking Time Seriously: Dynamic Regression.” *American Journal of Political Science* 52:184–200.

Enders, Walter. 2004. *Applied Econometric Time Series.* 2nd ed. Singapore: John Wiley and Sons.

Greene, William H. 2003. *Econometric Analysis.* 5th edition ed. New York: Prentice Hall.

Jackman, Simon. 2009. *Bayesian Analysis for the Social Sciences.* Wiley.

King, Gary, Michael Tomz & Jason Wittenberg. 2000. “Making the Most of Statistical Analyses: Improving Interpretation and Presentation.” *American Journal of Political Science* 44:347–361.